numero 190
giugno 1984
anno XVII
volume XXXII

LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

L'allattamento materno

Ha un valore contraccettivo di cui si tiene oggi sempre meno conto; in molti paesi il passaggio all'allattamento artificiale ha determinato un aumento della natalità e un peggioramento della salute dell'infanzia

di R. V.Short

L'allattamento materno è il contraccettivo della natura. Anticamente era conoscenza comune che una donna, durante l'allattamento, avesse scarse probabilità di rimanere incinta. Oggi il fatto che l'allattamento materno possa ancora svolgere un ruolo contraccettivo importante è diventato poco più che un mito nelle menti di molte persone. Tale cambiamento di mentalità ha avuto serie conseguenze sui tassi di crescita demografica e sulle condizioni di salute dell'infanzia, particolarmente in molti paesi in via di sviluppo.

Nei paesi sviluppati l'allattamento materno è visto generalmente come una conseguenza della riproduzione, un «extra» facoltativo che si può aggiungere alla fine, come i titoli di coda di un film, e che interessa solo pochi mentre difficilmente susciterà un'attenzione generale. Questo è un argomento che è stato ampiamente ignorato dalla professione medica. Non ci sono specialità mediche che se ne facciano carico. Il compito dell'ostetrico si conclude quando il bambino è stato partorito e la madre è stata dimessa dall'ospedale; il ginecologo a malapena considera la mammella come parte integrante dell'apparato della riproduzione; il pediatra usualmente si occupa più della cura dei bambini malati che di consigliare alla madre, in un normale ambiente domestico, l'allattamento al seno e pochi medici che si occupano di pianificazione familiare penserebbero di difendere l'allattamento materno per la sua efficacia contraccettiva.

Quindi è spesso il chirurgo generico che, per esclusione, è diventato lo specialista della mammella, poiché il cancro di questa è la neoplasia umana più comune. Per prevenire la malattia alcuni chirurghi arrivano al punto di raccomandare l'asportazione delle mammelle e la loro sostituzione con protesi al silicone. Non sorprende, dunque, che l'allattamento al seno goda di scarsa attenzione nei testi e nelle aule di medicina e quindi di poco interesse nella ricerca. La società sa molto di più sulla mammella della vacca da latte di quanto non sappia della tenera mammella umana.

L'incapacità da parte dei medici, degli scienziati, dei teologi, degli amministratori e dei politici dei paesi sviluppati di apprezzare il pieno significato dell'allattamento materno nel distanziare le nascite e nel salvaguardare la salute dei bambini ha avuto serie conseguenze nei paesi in via di sviluppo. Recenti stime dell'Organizzazione mondiale della sanità (OMS) suggeriscono che solo il 17 per cento circa delle coppie nei paesi in via di sviluppo usa forme moderne di contraccezione. Quindi l'83 per cento è totalmente dipendente dai controlli naturali della fecondità e, tra questi, l'allattamento al seno è di gran lunga il più importante.

La recente World Fertility Survey organizzata dall'International Statistical Institute di Londra ha mostrato che oggi, in molti paesi in via di sviluppo, le giovani donne istruite e benestanti, che vivono in città, stanno abbandonando l'allattamento al seno a favore di quello artificiale. Poiché esse sono prese a modello nelle loro comunità, la loro influenza si sta diffondendo alle donne di umili condizioni sia nelle città sia nelle aree rurali, donne che meno facilmente fanno uso di contraccettivi moderni e i cui bambini traggono il maggiore beneficio dall'allattamento materno.

Ron Lesthaeghe dell'Université Libre di Bruxelles ha calcolato che nel Bangladesh, dove solo il 9 per cento delle donne fa uso attualmente di contraccettivi e la durata media dell'amenorrea da lattazione (assenza di mestruazioni durante l'allattamento) è di 18 mesi e mezzo, l'uso di contraccettivi dovrebbe aumentare fino al 43 per cento, per mantenere la fertilità al suo livello attuale, se l'amenorrea da lattazione dovesse diminuire a sei mesi. Nel Kenya, dove solo il 7 per cento delle donne usa contraccettivi, buona parte del brusco recente aumento della fertilità fino al livello medio attuale che è di circa otto nati per donna, può essere attribuito direttamente a una diminuzione della durata dell'allattamento. Poiché è estremamente difficile, in questi paesi, provocare un rapido aumento nell'uso dei contraccettivi, è essenziale prevenire nel frattempo un'ulteriore riduzione dell'effetto contraccettivo dell'allattamento materno.

Sarebbe errato accusare i fabbricanti di latte in polvere di aver dato il via a questa tendenza ad abbandonare l'allattamento al seno: questa ha avuto inizio secoli fa. Tuttavia, essi sono in gran parte responsabili di perpetuarla e di favorirla oggi nei paesi in via di sviluppo, a danno della salute della madre e del bambino. Si spiega da sé il fatto che gli Stati Uniti, dove si trovano molte fabbriche di latte in polvere, siano stati l'unica nazione che, nel 1981, abbia rifiutato di firmare l'International Code of Marketing of Breast Milk Substitutes che, in accordo con le raccomandazioni dell'OMS, dovrebbe limitare la pubblicità aggressiva e la vendita di latte in polvere e di latte condensato.

Per comprendere il ruolo critico dell'allattamento materno nella riproduzione umana si dovrebbe affrontare l'argomento secondo una prospettiva evoluzionistica. Una delle chiavi del successo di *Homo sapiens* è il rallentamento del tasso di riproduzione della specie. Questo è derivato dalla posticipazione della pubertà fino al secondo decennio avanzato di vita (cosa sconosciuta negli altri mammiferi) e dal prolungamento dell'intervallo tra due nascite successive attraverso gli effetti contraccettivi dell'allattamento al seno.

Questi freni alla riproduzione hanno aperto nuovi orizzonti all'evoluzione umana. Lo sviluppo del voluminoso cervello umano, che è incompatibile con il passaggio del feto a termine attraverso un canale del parto stretto, deve avvenire dopo la nascita, rendendo così il bambino completamente dipendente dalla madre per un lungo periodo di tempo. Con la pubertà differita si è avuta un'infanzia più prolungata, che dà ai genitori la possibilità di impartire al bambino un gran numero di nozioni acquisite. È questa capacità di trasmettere informazioni, oltre ai geni, che ha dato alla specie umana la superiorità su tutti gli altri mammiferi.

È utile ricordare l'ambiente in cui si sono evolute queste limitazioni della riproduzione. La specie umana ha passato più del 90 per cento della sua esistenza a condurre una vita nomade cacciando e raccogliendo cibo. Questo modo di vivere è andato in gran parte perduto, ma rimane il fatto che, nei loro geni, gli esseri umani sono ancora cacciatori-raccoglitori; la «civiltà» è troppo recente per aver avuto una qualsiasi influenza apprezzabile sull'assetto genetico dell'uomo.

Le migliori indicazioni su questa esistenza scomparsa derivano da studi sulle poche società di cacciatori-raccoglitori di cibo rimaste in Africa, in Australia, in Nuova Guinea e nell'America Meridionale. Le popolazioni meglio studiate sono certamente i !Kung del deserto del Kalahari nell'Africa meridionale, che sono stati oggetto di ricerca da parte di Richard Borshay Lee dell'Università di Toronto e di Irven DeVore della Harvard University e dei loro studenti. Nancy Howell dell'Università di Toronto ha dimostrato che i !Kung, che non usano metodi contraccettivi moderni e non hanno pratiche di regolazione della fecondità come matrimoni tardivi, tabù sui rapporti sessuali durante l'allattamento o infanticidi, hanno una dimensione familiare completa media di 4,7 figli e un intervallo medio tra le nascite di 4,1 anni. Se si tiene conto della mortalità, si può calcolare che ci vorranno circa 300 anni per raddoppiare la popolazione.

Che questi lunghi intervalli tra le nascite siano ottenuti tramite l'allattamento al seno è evidente dal fatto che, se il bambino muore e la lattazione cessa, la madre sarà presto nuovamente gravida. Studi condotti da Melvin Konner di Harvard e da altri hanno mostrato che la madre tiene il bambino con sé ovunque vada e che dorme accanto a lui la notte. Il bambino si alimenta al seno frequentemente durante il giorno (circa quattro volte all'ora), ma ogni pasto dura solo un minuto o due; egli si alimenta frequentemente anche di notte, persino mentre la madre dorme. Se la donna partorisse più spesso di una volta ogni quattro anni, la sua libertà di movimento sarebbe severamente limitata e un'esistenza nomade diventerebbe per lei impossibile.

Benché questa alta frequenza delle poppate possa dapprima sembrare anormale, ci sono parecchi motivi per pensare che probabilmente essa rappresenta la norma nella specie umana. Un comportamento analogo è stato trovato tra i cacciatori-raccoglitori di cibo di Papua Nuova Guinea. Il tipo di comportamento sarebbe, inoltre, in armonia con la composizione biochimica del latte umano, che è povero di grassi, di proteine e di sostanza secca. Le specie che allattano i loro piccoli con minore frequenza (vacche, pecore, capre, cervi, conigli ecc.) hanno un latte più nutriente. I mammiferi più affini alla specie umana, gli scimpanzé e i gorilla, quando sono in libertà allattano i loro piccoli parecchie volte all'ora, producono un latte molto simile a quello umano, dormono di notte con il piccolo attaccato al petto e hanno intervalli tra le nascite di quattro o cinque anni.

Il meccanismo preciso tramite il quale l'allattamento inibisce la fecondità non è del tutto conosciuto. Certamente è cruciale lo stimolo della suzione in sé, non la produzione di latte. Esperimenti effettuati su pecore e canguri hanno dimostrato che, se il capezzolo viene denervato chirurgicamente, la lattazione continua normalmente, ma l'effetto contraccettivo della suzione va perduto.

I miei colleghi e io abbiamo potuto dimostrare che, nelle donne, la sensibilità del capezzolo aumenta spiccatamente al momento della nascita del bambino. Il fenomeno può essere importante per aumentare gli input inibitori al cervello. È oggi ben stabilito che l'atto della suzione provoca una liberazione riflessa dell'ormone ossitocina dalla parte posteriore dell'ipofisi; l'ormone causa la contrazione delle cellule mioepiteliali che circondano gli alveoli della ghiandola mammaria e quindi provoca l'emissione di latte. La suzione provoca anche una liberazione riflessa dell'ormone prolattina dalla parte anteriore dell'ipofisi; quest'ormone è ritenuto importante per mantenere la secrezione di latte a lungo termine. Si può dire, così, che l'ossitocina serve per il pasto di oggi, mentre la prolattina prepara quello di domani. Quanto più frequente è la suzione, tanto più alto è il livello di prolattina.

Le donne con tumori ipofisari secernenti prolattina generalmente non presentano ovulazione né mestruazioni e in esse la secrezione di ormone luteinizzante (un ormone della parte anteriore dell'ipofisi, che provoca l'ovulazione) risulta depressa. Se il tumore viene asportato chirurgicamente o la secrezione di prolattina viene inibita dal farmaco bromocriptina, la secrezione dell'ormone luteinizzante torna presto nella norma, seguita dalla ripresa dei cicli mestruali ovulatori.

Questa catena di eventi ha condotto all'opinione diffusa che l'alta produzione di prolattina durante la lattazione inibisca la secrezione dell'ormone luteinizzante e possa anche avere un effetto inibitore diretto sulle ovaie, con conseguente soppressione dell'ovulazione. Questo punto di vista è tuttora in discussione poiché il reso *(Macaca rhesus)*, in cui la secrezione di prolattina è stata soppressa con la bromocriptina, non avrà ovulazione se il piccolo continuerà a succhiare. Inoltre, le donne con amenorrea conseguente a un eccesso

Nel grafico sono riportati gli intervalli tra le nascite per tre differenti tipi di allattamento al seno in assenza di contraccettivi. Le donne dei !Kung *(a)*, un popolo di cacciatori-raccoglitori dell'Africa, allattano i loro piccoli di frequente per tre o quattro anni, raggiungendo un intervallo medio tra le nascite di 4,1 anni a una media di 4,7 nati vivi per tutto il periodo fecondo. Gli utteriti dell'America Settentrionale *(b)*, che allattano secondo schemi rigidi, cominciano a dare al bambino cibi integrativi pochi mesi dopo la nascita e lo svezzano entro un anno, hanno un intervallo medio tra le nascite di due anni e una media di 10,6 nati vivi. Una donna del Regno Unito *(c)*, che non aveva mai allattato i suoi bambini e la cui lattazione era stata soppressa dopo ciascun parto con stilbestrolo o etinilestradiolo, ha avuto un intervallo medio tra le nascite di 1,3 anni.

di prolattina ritornano a una normale secrezione di ormone luteinizzante quando sono trattate con naloxone, persino se il livello di prolattina rimane alto.

Un'ipotesi più attraente, per spiegare l'effetto contraccettivo dell'allattamento al seno, è quindi quella secondo la quale input nervosi, provenienti dal capezzolo, raggiungano l'ipotalamo. Qui essi hanno l'effetto di stimolare la liberazione di beta-endorfina, che si ritiene sia in grado di sopprimere la produzione dell'ormone di liberazione della gonadotropina da parte dell'ipotalamo. Ne risulta una diminuzione della secrezione dell'ormone luteinizzante e quindi una mancanza di ovulazione.

La durata dell'allattamento al seno nei paesi in via di sviluppo varia notevolmente. La posizione assegnata a ogni paese è una media; in generale le madri più giovani (dai 15 ai 24 anni) hanno allattato per un periodo più breve rispetto alle donne appartenenti ai gruppi di età di 25-34 e di 35-49 anni. I dati sono stati raccolti nel 1980 nel corso della World Fertility Survey, diretta dall'International Statistical Institute di Londra.

Nel Bangladesh le madri che vivono in città e quelle che vivono in campagna differiscono per quanto riguarda il tempo dedicato ad allattare al seno; in ogni coppia di istogrammi le prime sono indicate in colore e le seconde in grigio. In ogni gruppo di tre coppie di istogrammi, riferito a una determinata classe di età, la coppia a sinistra rappresenta le donne senza istruzione scolare, la coppia al centro le donne con una istruzione scolare elementare e la coppia a destra le donne che hanno terminato le scuole secondarie. In un certo numero di paesi in via di sviluppo le donne giovani, che vivono in città e che sono istruite, tendono ad abbandonare, o ad abbreviare, il periodo di allattamento al seno.

Gli stessi input provenienti dal capezzolo inibiscono anche la secrezione della dopammina da parte dell'ipotalamo e, poiché di norma la dopammina inibisce la secrezione di prolattina, quest'ultima aumenta immediatamente. Quindi il livello della prolattina è un utile indice del grado di inibizione provocato dalla suzione a livello dell'ipotalamo, ma può essere completamente non correlato con la causa dell'inibizione dell'ovulazione.

Gli effetti ormonali dell'allattamento al seno sono mossi dallo stimolo della suzione. Sono indicati qui alcuni ormoni prodotti dall'ipotalamo e dall'ipofisi. Quando l'effetto di un ormone consiste nello stimolare o incrementare un processo, esso viene rappresentato da una freccia continua. L'inibizione di un ormone o di un processo è rappresentata da una freccia tratteggiata. Poiché la secrezione di prolattina da parte dell'ipofisi è normalmente inibita dalla secrezione di dopammina da parte dell'ipotalamo, diminuendo i livelli di dopammina aumenta la secrezione di prolattina. L'ossitocina viene sintetizzata dall'ipotalamo e passa lungo le fibre nervose dirette alla parte posteriore dell'ipofisi, da cui è liberata nella circolazione sistemica, per facilitare l'emissione di latte.

Poiché, evidentemente, lo stimolo della suzione è il fattore cruciale nel provocare l'effetto contraccettivo, si dovrebbe esaminare in dettaglio il comportamento durante l'allattamento per comprendere ciò che avviene. Peter W. Howie e Alan S. McNeilly, un tempo miei colleghi all'unità del Medical Research Council di Edimburgo, hanno studiato 27 donne scozzesi che allattavano al seno per tutto il periodo della lattazione, controllando la loro attività ovarica tramite dosaggi degli estrogeni e del pregnandiolo nelle urine e registrando accuratamente la frequenza e la durata delle poppate. Come gruppo di controllo hanno seguito 10 donne che avevano scelto per i loro bambini l'allattamento artificiale fin dalla nascita.

I risultati sono stati molto interessanti. Mentre le donne che erano ricorse all'allattamento artificiale hanno cominciato a mestruare in media dopo otto settimane dal parto e a ovulare dopo 11 settimane, le madri che allattavano al seno hanno iniziato a mestruare in media dopo 33 settimane dal parto e a ovulare dopo 36 settimane. Nessuna donna ha ovulato se allattava al seno sei o più volte al giorno, e per più di 60 minuti al giorno, e nessuna donna ha ovulato durante questo tipo di allattamento non integrato artificialmente. Quando le donne sono state divise in due gruppi a seconda se avevano ovulato tardi (più di 40 settimane dopo il parto) o presto (meno di 30 settimane dal parto), si è potuto notare che il gruppo che aveva ovulato più tardi aveva allattato al seno più a lungo (durata media della lattazione 53 settimane), lo aveva fatto con una maggiore frequenza delle poppate, aveva prolungato la poppata notturna e aveva introdotto più gradualmente un'integrazione alimentare per il bambino. Quando, circa 16 settimane dopo il parto, sono stati introdotti gli alimenti integratori, c'è stato un rapido declino della frequenza delle poppate e del livello di prolattina nel sangue. Entro 16 settimane dall'introduzione di questo tipo di alimentazione, in 20 delle 27 donne, i follicoli ovarici avevano cominciato a svilupparsi e 14 di esse avevano già ovulato.

James B. Brown dell'Università di Melbourne ha condotto uno studio più ampio su 55 donne che allattavano al seno. Esse tendevano ad allattare per un tempo più lungo delle donne di Edimburgo, ma Brown non è riuscito a stabilire nei particolari questo tipo di comportamento. Combinando i risultati degli studi di Edimburgo e di Melbourne si possono trarre alcune conclusioni generali riguardo alla ripresa della fecondità nella donna dopo la nascita del bambino.

Nella grande maggioranza delle donne che allattano al seno le ovaie sono inattive per la maggior parte del periodo dell'amenorrea da lattazione. Con il trascorrere del tempo dopo la nascita, quando gradualmente nella dieta del bambino vengono introdotti alimenti integrativi, la frequenza e la durata delle poppate comincia a diminuire e l'attività ciclica delle ovaie riprende. Così, alla fine, molte donne che nutrono i loro figli al seno cominceranno a mestruare e a ovulare mentre in esse è ancora in atto la lattazione. Questo esempio spiega perché si ritenga comunemente che l'allattamento al seno sia una forma di contraccezione che non dà affidamento.

Negli studi di Edimburgo e di Melbourne, uno sguardo agli eventi endocrini che precedono la prima mestruazione dopo il parto rivela tuttavia che, nella maggioranza dei casi (58 per cento), le donne non ovulano prima di mestruare. La mestruazione, in questi casi, è presumibilmente il risultato di un calo di estrogeni, dovuto alla degenerazione di un follicolo che si stava sviluppando nell'ovaio. Quando davvero l'ovulazione ha preceduto la prima mestruazione, essa è stata seguita, di solito, da una fase luteinica del ciclo breve o insufficiente. Sembra improbabile che, in quelle donne, il corpo luteo dell'ovaio sarebbe riuscito a sostenere una gravidanza anche se un uovo fosse stato fecondato. Solo nel 19 per cento dei casi di Melbourne vi era un ciclo di ovulazione normale, precedente la prima mestruazione dopo il parto.

Poiché la massima probabilità di concepimento in un normale ciclo mestruale è soltanto del 25 per cento circa, il dato suggerisce che solo il cinque per cento delle donne che hanno rapporti sessuali non protetti durante l'allattamento prima della ricomparsa delle mestruazioni rimarrà incinta. Questo valore è in buon accordo con l'incidenza dei concepimenti (dal due all'8 per cento) durante l'amenorrea da lattazione, osservata nei paesi in via di sviluppo. A coloro che vivono in paesi sviluppati questo tasso di mancata contraccezione durante l'amenorrea da lattazione potrebbe sembrare troppo alto, ma bisogna ricordare che, nella situazione di un paese in via di sviluppo, la protezione fornita dal migliore dei contraccettivi moderni non è superiore a questa perché molte donne non usano i preparati in modo adeguato.

Un confronto tra le donne di Edimbur-

I cibi integrativi somministrati a un bambino allattato al seno dalla terza alla ventiquattresima settimana di vita hanno molteplici effetti. Qui i risultati sono centrati sul momento dell'inizio dell'integrazione alimentare *(linea verticale in colore)*. Sono riportati, dall'alto, gli effetti sul numero di poppate in un giorno, sul tempo totale di allattamento in un giorno, sul livello dell'ormone prolattina nella circolazione sistemica della madre e sull'attività delle ovaie, espressa come percentuale di donne con un incremento nella sintesi degli estrogeni da parte delle ovaie *(in nero)* e in ovulazione *(in colore)*. I dati sono ricavati da uno studio condotto a Edimburgo su 27 madri.

go e di Melbourne e quelle !Kung rivela molte differenze sorprendenti. Per la maggior parte le prime svezzavano il loro bambino entro un anno e mezzo di età, mentre le donne !Kung continuavano ad allattare il proprio per circa tre anni e mezzo. Poche donne di Edimburgo e di Melbourne allattavano il loro bambino più di otto volte nelle 24 ore; per le donne !Kung il numero delle poppate era circa 48 durante le 12 ore diurne. (Il tempo totale per giorno - tra i 100 e i 150 minuti - era probabilmente all'incirca lo stesso per tutte le donne.)

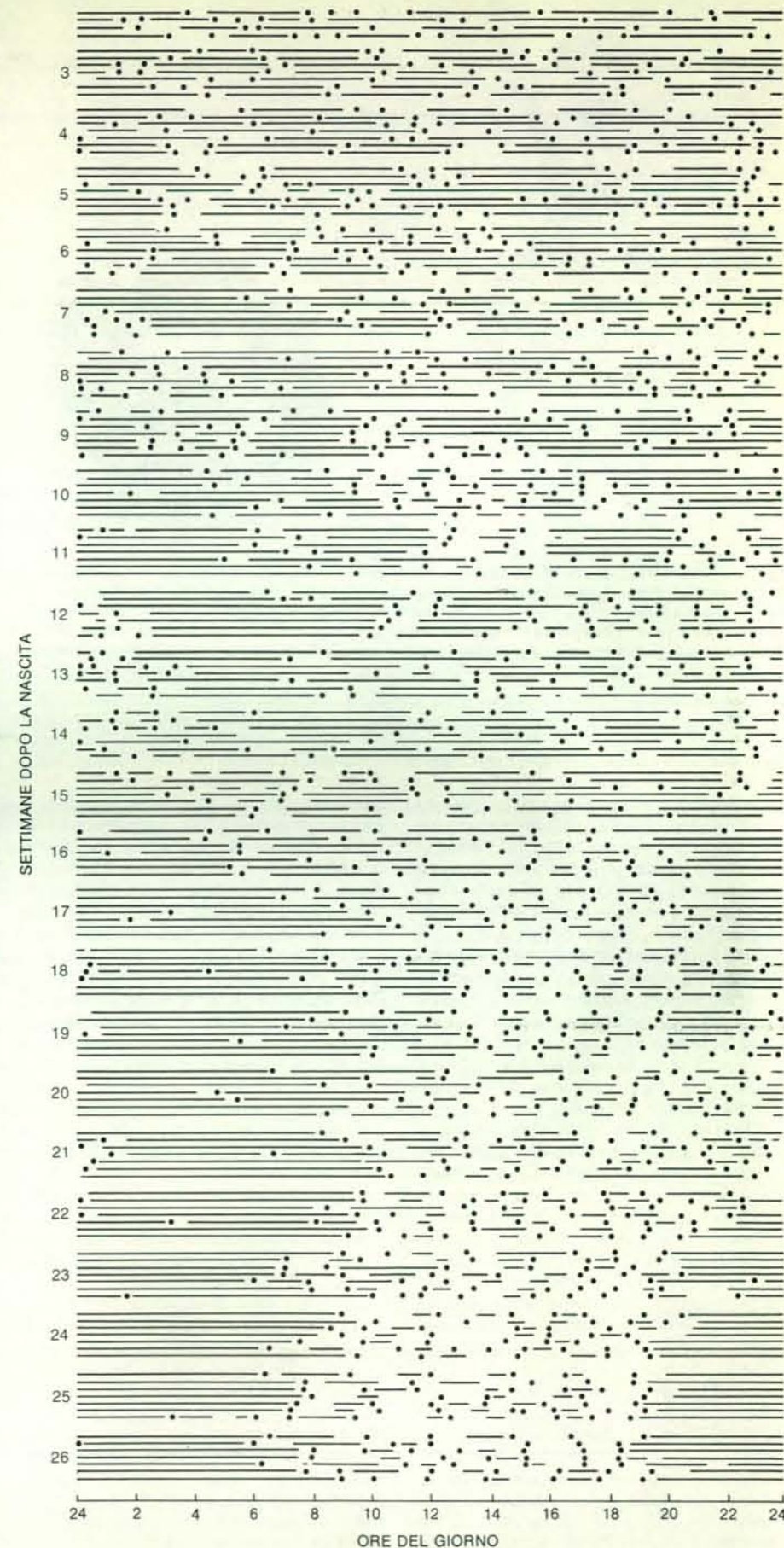
Una differenza particolarmente importante può essere quella che le donne di Edimburgo e di Melbourne tendevano ad abolire le poppate notturne il più presto possibile, mentre le donne !Kung dormivano con il loro bambino, così che l'allattamento notturno continuava per tutta la durata della lattazione. È noto da esperimenti con i bovini che l'intervallo di tempo tra le poppate può essere particolarmente importante per l'inibizione della ovulazione dopo il parto. Se si permette ai vitelli di poppare solo per 30 minuti al giorno, essi crescono normalmente, ma le vacche tornano quasi subito all'estro. Se, invece, si permette ai vitelli di poppare a volontà, le vacche mostrano un periodo prolungato di inibizione da lattazione dell'estro. Quindi l'abbandono dell'allattamento notturno da parte di molte donne può permettere all'ipotalamo di sfuggire all'inibizione da suzione, e riprende così l'attività delle ovaie.

Alcuni hanno tentato di spiegare i periodi prolungati di amenorrea da lattazione come l'effetto di un'insufficiente alimentazione. Questa argomentazione non potrebbe certamente essere applicata alle donne di Edimburgo e di Melbourne, che erano tutte ben nutrite. John Bongaarts del Population Council ha riesaminato i dati relativi a un certo numero di paesi in via di sviluppo e ha concluso che gradi moderati di malnutrizione cronica hanno solo un effetto minimo sulla durata dell'amenorrea da lattazione. Vi può essere, tuttavia, una sottile interazione tra livello di nutrizione e frequenza delle poppate. Andrew S. Loudon dell'Università di Edimburgo, studiando il cervo nobile in Scozia, ha dimostrato che le femmine che pascolano su terreni poveri di montagna producevano molto meno latte di quelle che pascolano su terreni perennemente erbosi. Come risultato i piccoli denutriti si alimentavano molto più frequentemente e le madri presentavano livelli di prolattina più elevati e ritornavano all'estro più tardi rispetto a quelle ben nutrite.

Roger G. Whitehead e i suoi colleghi della Dunn Nutrition Unit del Medical Research Council di Cambridge hanno condotto degli studi nel Gambia, dove alle donne denutrite che allattavano veniva fornita un'alimentazione integrativa (700 chilocalorie al giorno) sotto forma di biscotti altamente energetici; essi hanno concluso che la produzione di latte non aumentava, benché i livelli di prolattina diminuissero e l'ovulazione riprendesse più presto che nelle donne di controllo senza integrazione alimentare. Presa alla lettera, questa testimonianza suggerirebbe che, quando una madre che allatta al seno è molto malnutrita, il bambino ha bisogno di un'integrazione alimentare indipendente, poiché se anche la si nutre meglio, la sua produzione di latte non aumenterà.

Questo ragionamento è sicuramente sbagliato. Qualsiasi allevatore di bovine da latte, che ritenesse il livello di nutrizione senza effetto sulla produzione di latte, non farebbe buoni affari. Alla luce dei risultati ottenuti con il cervo nobile sembra più probabile che l'integrazione alimentare abbia di fatto aumentato la produzione di latte nelle donne del Gambia e che l'aumento si sia riflesso in una minore frequenza delle poppate. (Questa non è stata accuratamente registrata nello studio.) Ne sono conseguiti una caduta del livello di prolattina e un precoce ritorno all'ovulazione. Avendo la possibilità di integrare o la dieta della madre soltanto o le diete della madre e del bambino l'una indipendentemente dall'altra, la prima alternativa sarebbe chiaramente preferibile. Essa non solo permetterebbe alla madre di produrre l'alimento ideale per il suo bambino in maggiore quantità, ma preserverebbe almeno in parte l'effetto contraccettivo dell'allattamento. Ulteriori informazioni sul modo in cui l'integrazione alimentare influisce sulla produzione di latte sono necessarie al più presto, poiché dovrebbero costituire la base della filosofia grazie alla quale si offre aiuto alle popolazioni dei paesi sottosviluppati che soffrono la fame.

Quando si dovrebbero aggiungere cibi integrativi alla dieta del bambino? Un modo di accostarsi al problema consiste nel costruire curve «normali» di crescita per i bambini (quasi invariabilmente ottenute a partire da bambini allattati artificialmente secondo una formula che utilizza latte vaccino o latte umano pastorizzato) e nel determinare quando il tasso di crescita dei bambini allattati al seno comincia ad allontanarsi da questi valori normali. Un metodo alternativo è stato quello di valutare la produzione di latte in una donna che allatta e nel calcolare, oltre al valore calorico del latte umano, il fabbisogno energetico nel bambino che cresce. Con una simile impostazione, John C. Waterlow della London School of Hygiene and Tropical Medicine ha concluso di recente che il solo allattamento al seno non sarebbe in grado di soddisfare, dopo circa tre mesi, il fabbisogno energetico di molti bambini. D'altro canto, Chung H. Ahn e William C. MacLean della Johns Hopkins University School of Medicine hanno dimostrato che i bambini (figli di donne che avevano preso parte al programma La Lèche), allattati esclusivamente al seno per un periodo variabile da sei a 12 mesi, sono cresciuti in modo perfettamente normale, almeno per i primi nove o 10 mesi.

Un altro modo per determinare quan-

Sono qui riportati i cicli d'allattamento di un bambino nutrito al seno dall'undicesimo al centottantaduesimo giorno di vita. Le linee continue rappresentano i periodi di sonno, i puntini le poppate, gli spazi vuoti i periodi di veglia. È evidente che la madre ha abituato gradualmente il bambino a tralasciare il pasto notturno, minando così l'effetto contraccettivo dell'allattamento al seno.

Antichi recipienti usati per alimentare i bambini testimoniano che la tendenza ad abbandonare l'allattamento al seno è cominciata molto tempo fa. Questo vasellame fa parte della collezione del Wellcome Institute for the History of Medicine di Londra: il recipiente a sinistra proviene da Cipro e risale circa al X secolo a.C.; quello a destra proviene dall'Italia meridionale e risale al IV o al III secolo a.C. Un pezzetto di spugna posto nel beccuccio era forse utilizzato come tettarella.

do l'integrazione alimentare si rende necessaria è quello di osservare la bocca del bambino. Il momento in cui spuntano i denti venne sicuramente stabilito nel corso dell'evoluzione per fornire al bambino un mezzo alternativo per alimentarsi in un momento appropriato del suo sviluppo. I cacciatori-raccoglitori di cibo non avevano a disposizione per i loro figli cibi sofisticati; il bambino doveva essere svezzato direttamente con la dieta dell'adulto e quindi doveva saper masticare. Poiché gli incisivi non compaiono fino a circa sei mesi dalla nascita e i denti molari fino a nove o 10 mesi, se ne può dedurre che è questo il momento naturale per iniziare l'alimentazione integrata.

Un chiaro errore è la madornale sottovalutazione della potenziale produzione di latte umano. La cifra su cui si concorda in genere è di 700 millilitri al giorno quando il bambino ha sei settimane, mentre aumenta a circa 800 millilitri dal terzo mese di vita in poi. Peter E. Hartmann dell'Università della Western Australia ha però dimostrato che le donne australiane che allattano i bambini esclusivamente al seno, se questi lo desiderano, per almeno 15 mesi possono facilmente produrre 1500 millilitri di latte al giorno. Con i gemelli la produzione può essere di oltre 2500 millilitri. La ghiandola mammaria è quindi uno splendido organo domanda--offerta: più spesso il latte viene rimosso da essa, più latte sarà prodotto. Il fenomeno è probabilmente il risultato di un meccanismo intraalveolare, per mezzo del quale una dilatazione dell'alveolo, causata dalla presenza del latte, inibisce l'ulteriore sintesi di latte; il meccanismo preciso non è però noto.

In molti paesi in via di sviluppo e industrializzati è assai diffusa la pratica che la madre cominci molto presto a nutrire il bambino con integrazioni artificiali, a volte fin dal giorno della nascita. Questa pratica può servire solo a diminuire la produzione di latte materno, a esporre il bambino a un rischio non necessario di infezioni e a rendere inefficace l'effetto contraccettivo dell'allattamento al seno. Vi è anche qualche nuova prova secondo cui l'introduzione di una integrazione alimentare nei primi quattro mesi di vita può facilitare il manifestarsi dell'eczema infantile. Sarebbe di gran lunga più sensato incoraggiare la madre ad allattare il suo bambino esclusivamente al seno, almeno per i primi sei mesi, e arricchire in questo periodo la sua dieta anziché quella del bambino.

Un'altra controversia corrente riguarda il tipo di contraccettivo che una madre che sta allattando dovrebbe usare e il momento in cui dovrebbe cominciare a usarlo. Nei paesi sviluppati, dove le donne generalmente allattano al seno per un periodo di tempo relativamente breve e dove ritengono garantita la disponibilità di un metodo contraccettivo quasi perfetto, la pratica è stata di consigliare metodi barriera, dispositivi intrauterini o la minipillola di soli progestinici, entro un mese circa dal parto. (La pillola anticoncezionale standard contiene una combinazione di estrogeni e di progestinici; gli estrogeni non sono presenti nella minipillola perché inibiscono la lattazione.)

Nei paesi in via di sviluppo, dove le donne allattano molto più a lungo e dove i metodi anticoncezionali sono scarsi e poco usati, la situazione è completamente diversa. Come ho messo in rilievo, la protezione contraccettiva fornita dall'amenorrea da lattazione è almeno efficace quanto qualsiasi anticoncezionale moderno. Poiché l'amenorrea da lattazione può durare per un anno e più, sembra poco motivata una «doppia copertura» in questo periodo. In effetti, uno studio effettuato nel Bangladesh ha suggerito che la somministrazione della pillola estroprogestinica alle donne che allattano può far realmente aumentare la loro fecondità, dato che il suo effetto depressivo sulla lattazione probabilmente affretta lo svezzamento. Durante lo svezzamento un gran numero di donne non prenderà la pillola e perderà così ogni protezione contro la gravidanza.

Uno dei più efficaci e innocui contraccettivi per le donne che allattano nei paesi in via di sviluppo e il più semplice da somministrare è il Depo-Provera. Si tratta di un progestinico somministrato per iniezione una volta ogni tre mesi, periodo durante il quale sopprime l'ovulazione e la mestruazione. Teoricamente questo contracettivo dovrebbe essere somministrato verso la fine della lattazione, proprio prima del momento in cui si prevede la ripresa del ciclo mestruale.

Affronto qui un argomento controverso. Studi condotti dall'Organizzazione mondiale della sanità non hanno ancora potuto identificare un rischio per la salute associato all'uso del Depo-Provera e non hanno dimostrato effetti negativi sulla produzione e sulla composizione del latte materno. D'altro canto, molte autorità nei paesi sviluppati si oppongono all'uso del Depo-Provera, in quanto non è ancora stato dimostrato in modo conclusivo che esso sia privo di effetti collaterali. A causa dell'opposizione mossa da associazioni di consumatori, la US Food and Drug Administration ha fatto marcia indietro sulla decisione presa di approvare il farmaco come contraccettivo negli Stati Uniti. Il caso è in appello. Pertanto il Depo-Provera non può entrare nei programmi di pianificazione delle nascite condotti dalla US Agency for International Development.

Il ministro inglese della Sanità ha respinto, per «ragioni non mediche» non meglio precisate, il parere del suo Committee on the Safety of Medicines e ha rifiutato di approvare il Depo-Provera come contraccettivo. Il Comitato nazionale svedese per la salute pubblica e l'assistenza lo ha approvato come contraccettivo utilizzabile nell'ambito del paese, ma le autorità competenti per lo sviluppo internazionale hanno ritirato tutti i fondi per la sua distribuzione a livello mondiale. Il ministero della Sanità della Repubblica Federale Tedesca ha recentemente approvato la somministrazione del Depo--Provera per le donne che allattano, revocando una decisione precedente che non permetteva l'uso di questo farmaco perché comportava un «rischio non giustificabile per la salute». L'Australian Drug Evaluation Committee ha bandito l'uso del Depo-Provera come contraccettivo per «mancanza di dati clinici», nonostante che circa un milione e mezzo di donne in tutto il mondo lo stia utilizzando come misura anticoncezionale.

Una delle ragioni che sono alla base di questa esitazione si collega al rischio ancora sconosciuto, derivante dalle piccole quantità di progestinici che vengono eliminate nel latte materno e che sono assorbite dal bambino. Ma anche i bambini allattati artificialmente e che ricevono latte vaccino o preparati a base di quest'ultimo sono esposti a piccole quantità di steroidi. Il fatto è che molte vacche che producono latte sono gravide e il latte è una delle vie di escrezione per il progesterone e il solfato di estrone ad alti livelli, associati alla gestazione. Quando una donna che sta allattando rimane incinta, anche il suo latte contiene progesterone ed estrogeni - una condizione che il Depo-Provera è specificamente designato a prevenire. Sarebbe stato quindi meglio per tutti gli interessati se le istituzioni preposte al controllo dei farmaci avessero approvato il Depo-Provera come contraccettivo per le donne che allattano e avessero bandito, invece, l'allattamento dei bambini con tutti i prodotti derivati dal latte vaccino.

Una ragione per la quale, nei paesi in via di sviluppo, un continuo apporto di latte materno è di vitale importanza per la salute del bambino riguarda il passaggio di globuli bianchi dall'intestino alla mammella. Questa via è stata scoperta solo nello scorso decennio. Le placche di Peyer (grandi masse di tessuto linfoide associate all'apparato digerente e situate nell'intestino tenue) contengono cellule designate al riconoscimento degli antigeni e sensibilizzate agli agenti patogeni che nel corso della sua esistenza la madre può aver ospitato nell'intestino. Se, mentre allatta, la madre ingerisce con il cibo uno di questi agenti, le cellule sensibilizzate migrano dalle placche di Peyer nella circolazione sistemica e da qui raggiungono le mammelle, diventando plasmacellule che liberano nel latte una specifica immunoglobulina A, diretta contro quel determinato agente patogeno. Il processo può essere facilitato dall'alto livello di prolattina in circolo.

Durante il quarto mese di lattazione, una donna può secernere nel latte fino a 0,5 grammi di anticorpi al giorno. Gli anticorpi non sono assorbiti dal bambino, ma rimangono nell'apparato digerente dove svolgono una funzione di vitale importanza nel prevenire le infezioni gastrointestinali (la causa principale di mortalità infantile). Nessun tipo di latte artificiale, per quanto accuratamente studiato, può sostituire questa protezione immunologica su misura, fornita dal latte umano.

Il costante declino dell'allattamento al seno è una delle principali tragedie per l'umanità. Esso ha causato, allo stesso tempo, un incremento eccessivo del numero di nascite e un enorme aumento della mortalità infantile. Se ne sa già abbastanza per prevenire il ripetersi di questa tragedia nei paesi in via di sviluppo. Il compito è ora quello di tradurre la consapevolezza in azione.

L'etica mafiosa e lo spirito del capitalismo

La grande trasformazione postbellica della società italiana ha avuto effetti rilevanti sul potere e sul comportamento della mafia mutando il suo ruolo di mediazione in una attività di tipo imprenditoriale

di Pino Arlacchi

Pochi argomenti come quelli associati con la parola «mafia» colpiscono la fantasia del grande pubblico in quasi tutti i paesi del mondo. L'idea di un potere misterioso e pervasivo, strutturato come una moderna società multinazionale, ma con radici storiche e culturali molto antiche, capace di sfidare con successo forze dell'ordine e autorità ufficiali, è così profondamente radicata nella mentalità collettiva da rendere molto arduo il compito di suggerire una immagine differente del fenomeno mafioso.

In questo articolo mi propongo di esporre alcuni dei principali risultati delle ricerche che ho condotto negli ultimi cinque anni sul problema della mafia nell'Italia meridionale. Tali risultati smentiscono alcuni degli stereotipi popolari più diffusi sull'argomento, e contraddicono anche diversi luoghi comuni imperanti nella comunità scientifica a proposito della struttura e del *modus operandi* del potere mafioso.

Dividerò l'esposizione in tre parti. Nella prima tenterò di delineare un «tipo ideale» di mafia tradizionale, valido per il periodo di tempo che va dall'unificazione italiana alla fine della seconda guerra mondiale, cercando anche di definire alcune delle relazioni vigenti tra il fenomeno mafioso e l'ambiente socioeconomico che lo ha espresso. Nella seconda parte descriverò i cambiamenti di tale fenomeno durante la «grande trasformazione» dell'Italia e del Mezzogiorno negli anni cinquanta e sessanta. Esporrò infine il concetto di «mafia imprenditrice» tentando di verificarne la validità conoscitiva nell'analisi di alcuni importanti problemi politici ed economici del Mezzogiorno contemporaneo.

Il tipo ideale della mafia tradizionale

Fino a qualche decennio fa, la maggioranza della popolazione della Sicilia occidentale intendeva indicare con il termine «mafia» due insiemi di fatti, due fenomeni sociali che - pur essendo strettamente connessi nella vita pratica - devono essere tenuti distinti sul piano analitico. Il primo significato della parola indicava una particolare forma di comportamento e di potere largamente diffusi in determinati strati della società. Il secondo si riferiva invece a un insieme di piccoli gruppi parentali-amicali autonomi che agivano in rapporti di cooperazione e di lotta per il controllo di un dato territorio e delle sue risorse.

«Comportarsi in maniera mafiosa» significava «farsi rispettare», essere «uomini d'onore», capaci di vendicare con le proprie forze qualunque offesa arrecata alla propria personalità o alle sue estensioni, e capaci di arrecare qualunque offesa a un nemico. Tale comportamento, sia che fosse di natura difensiva o aggressiva, veniva non solo giustificato ma incoraggiato e idealizzato dalla particolare cultura in cui viveva il mafioso, anche se il ricorso alla violenza si scontrava con il diritto statale. Anzi, una parte significativa del prestigio-potere conferito dall'azione mafiosa derivava proprio dal fatto che essa costituiva un'aperta infrazione delle norme e delle istituzioni giuridiche ufficiali.

Nei sistemi culturali di tipo mafioso, la forza personale dell'individuo contava più immediatamente e ovviamente che altrove nel determinare la distribuzione dell'onore tra i diversi membri della società. Né la nascita né le istituzioni ufficiali avevano una influenza determinante sulla forma di questa distribuzione. Uomini d'onore si diventava, non si nasceva. La competizione per l'onore era libera e chiunque poteva entrare in gara.

La competizione per l'onore non veniva accuratamente delimitata e istituzionalizzata, come avviene nelle moderne forme di conflitto sportivo, scolastico, mercantile e nella lotta di classe e di gruppo. Tutti i mezzi erano buoni, e venivano perciò trasferite nella lotta tra uomo e uomo le forme più arcaiche del conflitto sociale: la rapina, la devastazione, il sequestro e la strage. Se si osserva la carriera dei principali mafiosi si resta colpiti dal numero e dalla gravità dei conflitti da cui essi sono usciti vittoriosi, specie durante i primi tempi del loro tirocinio.

In un sistema fondato sulla lotta priva di regole per la supremazia, non esiste un modo più definitivo di affermare la propria eccellenza del togliere la vita di un altro uomo. L'assassinio di Bernardo Cassini descritto dall'antropologo A. Blok nel suo studio *The Mafia of a Sicilian Village, 1860-1960*, procura al suo uccisore non solo terra e potere, ma anche il titolo onorifico di *don*. Molta parte del prestigio del giovane mafioso nasceva dall'omicidio. Un pastore o un bracciante qualsiasi poteva conquistare una reputazione di violenza e diventare un uomo con cui bisognava fare i conti.

Ma il fenomeno mafioso tradizionale non consisteva solo di un comportamento violento e individualistico, messo in atto allo scopo di conquistare una supremazia che era fonte di stima e di ammirazione popolare. Esisteva anche un secondo aspetto, che aveva una importanza pari a quella del primo e un segno radicalmente opposto. Questo secondo momento può essere descritto nei termini di un processo di istituzionalizzazione dell'onore e di trasformazione del medesimo in un potere riconosciuto come legittimo. Quando la competizione per la supremazia si risolveva con la vittoria di un concorrente molto forte, accadeva che quest'ultimo cercasse di stabilizzare la propria posizione istituendo un monopolio territoriale della violenza molto simile a quello statale. Il mafioso adesso tentava di «congelare» l'esistente distribuzione dell'onore e di attenuare, regolamentandola e controllandola, la lotta di tutti contro tutti che stava alla base del sistema. Per raggiungere questi scopi, egli non doveva più presentarsi come il campione spregiudicato e feroce, capace di annientare qualunque avversario, ma come un uomo di governo saggio e prudente, alla portata di tutti, disposto a proteggere, mediare, consigliare e reprimere a seconda delle circostanze e sempre in vista dell'ordine e del bene collettivi.

Affrontando il problema delle matrici del fenomeno mafioso con il metodo della sociologia comparata, sono emersi come tratti caratteristici dell'ambiente che «produce» la mafia la conflittualità e l'instabilità sociale e la modernità della struttura economica. Si è potuto così isolare come sede privilegiata della fenomenologia mafiosa classica uno specifico sistema socioeconomico contrassegnato da un endemico contrasto tra la struttura mercantile-capitalistica della sua economia e il tradizionalismo della sua struttura culturale. Il disegno a china è di Lida Barazzutti.

La ricerca di un principio di regolazione del conflitto era d'altra parte molto viva nelle zone di mafia. Il sistema socioeconomico era qui sottoposto a una continua, reale minaccia di disintegrazione ed esisteva perciò un forte bisogno di un potere sovraindividuale, pubblico, in grado di creare anche una parvenza di ordine collettivo. L'incontro tra questa esigenza di autoregolazione del sistema e la necessità di conservazione dell'onore da parte degli uomini di rispetto faceva sì che questi ultimi venissero investiti di importanti funzioni pubbliche di salvaguardia delle norme tradizionali e di mantenimento dell'ordine costituito. I mafiosi diventavano così giudici civili e penali, ordinatori, mediatori, riassumendo nella propria persona molti delicati poteri normalmente esercitati dallo stato.

Tra le molteplici funzioni esercitate dal potere mafioso nella società tradizionale un ruolo di particolare rilevanza spettava alle funzioni di «mediazione» dei conflitti interni alla comunità locale e dei rapporti tra questa e il mondo esterno. Le relazioni con le grandi agenzie esterne della vita economica e politica, con la città, con lo stato e con il mercato nazionale e internazionale venivano favorite dalla professione ufficiale della grande maggioranza degli uomini d'onore affermati. Essi erano molto spesso commercianti di grano, vino, olio, agrumi, macellai, allevatori di bestiame, proprietari e affittuari agricoli.

È importante tenere presente come tutte le fondamentali attività del mafioso tradizionale non si svolgessero, in tempi

normali, in opposizione frontale alle istituzioni ufficiali. Mafiosi e organi dello stato - teoricamente antagonisti in quanto competitori per il monopolio territoriale della forza - collaboravano con metodi spesso simili nella repressione delle minacce all'ordine costituito. In molti episodi di banditismo e di criminalità comune, come in molti episodi di dissenso politico organizzato e di lotta di classe, l'intervento del potere armato dei mafiosi a fianco delle forze dell'ordine ufficiali è stato determinante per la cattura, l'uccisione o la neutralizzazione dei «devianti».

Dal 1860 in poi - a eccezione del periodo fascista - è esistito un sostanziale riconoscimento dell'autorità mafiosa da parte dell'autorità statale. Lo stato italiano concesse ai mafiosi un'ampia delega per l'esercizio delle funzioni di ordine pubblico in larghe aree delle estreme regioni meridionali. Tale delega produceva anche un importante effetto di integrazione politica. Numerosi segmenti politici locali entravano a far parte - tramite i mafiosi - del sistema politico nazionale. Il capomafia era quasi sempre un grande elettore locale il cui appoggio era necessario al candidato nelle elezioni politiche e amministrative. Molti uomini di governo siciliani - tra cui Vittorio Emanuele Orlando, che arrivò a diventare Presidente del Consiglio durante la prima guerra mondiale - hanno dovuto buona parte del loro successo elettorale ai loro legami con i gruppi mafiosi. In cambio del loro appoggio, i mafiosi ricevevano dagli uomini politici al governo favori per se stessi e per il proprio gruppo amicale-parentale, cioè per la propria cosca.

Siamo arrivati a questo punto all'analisi del secondo significato fondamentale del termine «mafia» nella società tradizionale, significato che è più vicino a quello odierno. Contrariamente a quanto viene suggerito da gran parte della pubblicistica letteraria e giornalistica sull'argomento, non esiste e non è mai esistita una organizzazione criminale segreta, gerarchica e centralizzata, come un «capo dei capi» residente in Sicilia o altrove, chiamata «mafia», «ndrangheta» od «onorata società», i cui membri siano legati l'un l'altro da giuramenti di mutua fedeltà e assistenza effettuati nel corso di tenebrose cerimonie. Abbiamo finora cercato di mostrare come il termine «mafia» indicasse piuttosto una particolare forma di comportamento e di potere, rinvenibile nelle azioni di una categoria di soggetti detti «mafiosi», «uomini d'onore», o, in Calabria, «ndranghetisti».

Dobbiamo ora sottolineare come l'attività di questi soggetti non si svolgesse in forme soltanto individuali, ma rappresentasse il prodotto della collaborazione di un gruppo più o meno ampio di consanguinei, di amici e di clienti chiamato «cosca». La cosca mafiosa tradizionale era un organismo semplice, ma solido, che non aveva nulla di formale e di burocratico. Al suo interno non esistevano statuti, né riti di iniziazione né tribunali: «Non ci sono presidenti, né segretari eletti in una forma qualsiasi, né ruoli dei soci. Il sodalizio è diretto e sfruttato quasi sempre da tre, quattro o cinque persone più autorevoli per l'età, l'intelligenza, la posizione sociale, le prove fatte, le condanne riportate e soprattutto per l'esperienza e la perizia maggiore nella difficile arte di delinquere impunemente», scriveva Gaetano Mosca all'inizio di questo secolo a proposito della tipica cosca mafiosa siciliana.

L'estensione spaziale controllata da ciascuna cosca era direttamente proporzionale al potere da essa detenuto e al prestigio attribuito al suo capo o ai suoi capi. La potenza di ciascuna cosca dipendeva a sua volta dal numero dei membri validi che la componevano. Nell'arco temporale di validità del nostro tipo ideale della mafia, e cioè tra il 1860 e il 1950, il numero dei membri interni effettivi di ciascuna cosca risultava essere alquanto limitato, non superando quasi mai le 15-20 unità. L'ampiezza di una cosca difficilmente oltrepassava tale soglia, a causa della natura dei rapporti che legavano i suoi membri. Il governo del gruppo diventava a un certo punto impossibile a causa dei conflitti interni e delle scissioni originate dalla formazione di nuove unità parentali-amicali autonome all'interno della casa-madre, e a causa del notevole grado di fluidità che contrassegnava i cerchi più esterni della cosca medesima. I grandi spazi fisici ed economici come una città, una provincia o una regione, non erano perciò mai controllati da un unico gruppo, ma da una serie di gruppi in relazioni di alleanza precaria o anche di aperto conflitto.

L'abitudine prevalente nelle zone mafiose di giudicare fatti, eventi e persone dal punto di vista dell'onore conseguito tramite la vittoria nelle lotte e nelle competizioni faceva sì che l'aggressività e la violenza venissero positivamente sanzionate. La partecipazione alla vita della comunità da parte delle categorie meno predisposte al conflitto per ragioni biologiche e culturali, come le donne, i bambini, i vecchi, risultava severamente limitata, a meno che i membri di questi gruppi non assumessero le vesti virili e bellicose delle categorie superiori. Il fascino delle donne di Bagnara - un paese della Calabria tirrenica in cui era presente una comunità di pescatori organizzata su base matriarcale - nasceva, agli occhi delle popolazioni circostanti e a quelli di Francesco Perri, autore del romanzo *Emigranti* (Lerici, Cosenza), dal fatto che «si diceva che portassero i rasoi nei capelli e maneggiassero il coltello più arditamente degli uomini». Le due immagini, di Enzo Sellerio *(a sinistra)* e di Informazione fotografica *(qui sopra)*, sono tratte, come quella di pagina 25, dal volume *Mafia: il fatto, la foto* (Idea editions, Milano), e sono pubblicate per gentile concessione dell'editore.

Onore e ricchezza

Il comportamento e il potere mafioso, nonché l'esercizio delle attività legali e illegali intraprese dai membri delle cosche mafiose, hanno finito per creare uno standard sociologico abbastanza preciso della figura del mafioso. I caratteri di questo standard sociologico sono l'origine popolare, l'appartenenza al ceto medio e il possesso di un potere territoriale precisamente delimitato.

I più importanti mafiosi siciliani e calabresi di questo secolo provengono tutti dagli strati più bassi della società. Vito Cascio Ferro e Calogero Vizzini erano figli di contadini poveri e contadini essi stessi all'inizio della loro carriera. Mommo Piromalli faceva il bovaro prima di diventare uno dei più temibili commercianti della Piana di Gioia Tauro. Il loro successo come uomini d'onore ha implicato il raggiungimento della ricchezza per loro stessi, per la loro famiglia e per la loro cosca. Sia nella valutazione popolare sia in quella del mafioso tradizionale, però, onore e ricchezza si accompagnavano, non si identificavano. La fondamentalità dell'onore nella struttura dell'azione mafiosa faceva sì che il movente economico puro e semplice - inteso sia nei termini della volgare sete di guadagno sia in quelli della religione dell'accumulazione - non riuscisse ad affermarsi come il supremo regolatore dei rapporti e delle posizioni sociali.

L'accumulazione, e ancor più la concentrazione, della ricchezza da parte dei mafiosi tradizionali, raggiunto un certo vago e un po' convenzionale livello, rallentava e si fermava. Oltre un certo limite, infatti, la ricchezza (e il potere a essa associato) diventavano per il mafioso ingombranti, difficili da difendere e da giustificare.

Il possesso della terra poteva essere il prodotto di una buona posizione nella gerarchia dell'onore e fonte di onore esso stesso. Ma la «scala» di questo possesso non superava di regola una certa soglia, oltre la quale la posizione onorifica del mafioso-proprietario diventava difficile da mantenere. Per fronteggiare con sicurezza l'aggressività dei numerosi sfidanti sarebbe occorso che egli si munisse di un apparato di guardie armate stipendiate, si alleasse strettamente all'élite del potere regionale e nazionale e rinunciasse a gran parte delle sue funzioni di mediazione all'interno del suo universo territoriale e culturale. Tutte e tre queste condizioni-conseguenze dell'accumulazione terriera erano - nella specifica situazione delle aree mafiose - antionorifiche in alta misura e comportavano l'abbandono di uno stile di vita gratificante nonché una sostituzione di valori che il mafioso tradizionale non era incentivato a compiere.

Questa conversione imperfetta tra onore e ricchezza influenzava notevolmente fisionomia e proporzioni del potere mafioso. Questo si presentava quasi sempre, infatti, come un potere concreto, locale, di villaggio, di quartiere, di zona, e mai nelle vesti di un potere impersonale astratto, di tipo regionale e nazionale. Vito Cascio Ferro, Calogero Vizzini e Giuseppe Genco Russo hanno vissuto per gran parte della loro vita in luoghi come Bisacquino (8400 abitanti nel 1951), Villalba (4900 abitanti alla stessa data), Mussomeli (15 800 abitanti). La consistenza dei loro patrimoni non superò mai i limiti di una fortuna di piccole e medie dimensioni, e l'orizzonte delle loro ambizioni rimase sempre proporzionato alla loro posizione economica.

Le matrici del fenomeno mafioso

L'elemento costituito dalle basi onorifico-competitive del fenomeno mafioso è quasi del tutto assente dagli studi sulla mafia tradizionale, fondati in massima parte su una visione della società siciliana che accentua gli elementi di staticità e di stratificazione presenti nella struttura di quest'ultima. I concetti cui fa riferi-

È nelle origini onorifico-competitive del potere mafioso che va cercata la spiegazione del fenomeno, apparentemente strano, del «trasferimento» di tale autorità nelle persone degli antagonisti al potere mafioso stesso. L'iniziale successo della lotta antimafia intrapresa in Sicilia dal prefetto Cesare Mori nel 1924 fu dovuto in buona parte all'intuizione da parte del prefetto delle radici «hobbesiane» del potere mafioso: «Se i siciliani hanno paura dei mafiosi, li convincerò che io sono il mafioso più forte di tutti», dichiarò ai suoi collaboratori. Cardine di tutta la sua strategia fu l'instaurazione di una competizione per l'onore con i mafiosi siciliani da svolgere in parte sul loro stesso terreno, in parte fidando sulla superiorità militare e organizzativa dello stato. La fotografia, del Centro documentazione Mondadori, ritrae Mori in tenuta da caccia.

mento la maggioranza degli studiosi a proposito della natura dei rapporti di proprietà e di produzione dominanti nelle aree mafiose della Sicilia tradizionale sono quelli di «feudalesimo», «feudo», «residuo feudale» e, soprattutto, «latifondo». Tali concetti evocano un universo caratterizzato da deboli flussi della mobilità verticale e da una distribuzione prefissata dell'onore e del potere tra i membri della comunità locale.

L'impostazione qui adottata sottolinea, al contrario, gli elementi di conflittualità e di instabilità sociale e la modernità della struttura economica come tratti caratteristici dell'ambiente che «produce» il fenomeno mafioso. L'individuazione di tali elementi è avvenuta nel corso della mia precedente ricerca sulle «strutture elementari del sottosviluppo» del Mezzogiorno tradizionale. L'uso di un metodo di sociologia comparata mi aveva consentito di isolare uno specifico sistema socioeconomico - la «società di transizione permanente», rinvenibile in forma concentrata nella Piana di Gioia Tauro, in Calabria, e in alcune aree della Sicilia occidentale pre-anni cinquanta, contrassegnata da un endemico contrasto tra la struttura mercantile-capitalistica della sua economia e il tradizionalismo della sua struttura culturale - come sede privilegiata della fenomenologia mafiosa classica. In questo tipo di società il comportamento mafioso costituiva la forma basilare dell'azione sociale e il potere mafioso aveva un ruolo decisivo nel processo dell'integrazione socioeconomica.

La società di transizione permanente era profondamente integrata nel mercato nazionale e internazionale delle merci agricole. La sua ossatura economica era costituita da una fascia ampia e instabile di medie imprese agricole e commerciali in perpetuo conflitto reciproco. Data l'altissima instabilità della vicenda economica, soggetta per ragioni sia tecnologico-produttive sia di andamento dei prezzi dei prodotti a fluttuazioni improvvise e imprevedibili, i flussi della mobilità ascendente e discendente avevano in questo tipo di società delle proporzioni così massicce da impedire la fissazione di una stabile impalcatura di classi sociali.

La mobilità generalizzata prodotta dalla cronica instabilità delle forze di mercato non era però a circuito chiuso. Non era inserita, cioè, in un movimento più ampio di costituzione di un nuovo sistema di rapporti sociali, in una tendenza di lungo o di medio periodo verso quella che Ferdinand Tönnies chiamava la *Gesellschaft*. Nella società di transizione permanente non venivano create nuove figure sociali. La società si ristrutturava periodicamente rimanendo in fondo inalterata, «avvitandosi» continuamente su se stessa. Questo andamento «a corsi e ricorsi» dava luogo a una situazione di «né *Gemeinschaft* né *Gesellschaft*», cioè, appunto, di transizione permanente.

Il potere mafioso veniva così a costituire, in questo sistema sociale, un originale quanto precario dispositivo di reintegro della coesione sociale minata dalle forze di mercato. Caratteristica essenziale della mafia tradizionale era infatti quella di essere non una organizzazione o un movimento con scopi e programmi specifici, ma il punto di confluenza di una serie di tendenze contrapposte: la difesa dell'intera società contro le minacce al suo tradizionale modo di vivere; le aspirazioni dei vari gruppi che la componevano a un «congelamento» di quel movimento ondulatorio della mobilità che ne distruggeva ogni istituzione e ogni sicurezza; le ambizioni personali e le aspirazioni di individui attivi e spregiudicati.

Dall'analisi fin qui svolta è possibile astrarre alcune precondizioni logico-strutturali di base, la cui presenza contemporanea potrebbe essere considerata necessaria allo sviluppo pieno del fenomeno mafioso tradizionale: *a*) la presenza determinante del mercato concorrenziale; *b*) una intensa conflittualità orizzontale tra individui e gruppi all'interno di un sistema culturale tradizionale; *c*) la debolezza o l'assenza di un principio di regolazione centralizzata dei rapporti economici e sociali o di un controllo statale della violenza.

La presenza del mercato concorrenziale costituisce un presupposto fondamentale del fenomeno mafioso poiché garantisce quella mobilità, instabilità e stratificazione di tutte le posizioni che incoraggiano enormemente la competizione interna. L'esistenza di una lotta generalizzata tra unità sociali elementari è in parte una conseguenza della penetrazione mercantile nella società, ma la guerra di tutti contro tutti si produce solo quando l'impatto tra il mercato e la cultura tradizionale non si risolve a favore né dell'uno né dell'altra. L'ultima precondizione esprime la necessità di uno stato di disgregazione economica e sociopolitica molto accentuata perché il potere mafioso possa affermarsi come forma dominante dell'integrazione. La presenza di grandi organizzazioni politiche ed economiche come gli stati nazionali, le imprese, i partiti e le burocrazie di massa, costituisce un elemento di coordinamento *ex-ante* dei rapporti sociali che «si oppone» alla regolazione *ex-post* effettuata dall'autorità mafiosa.

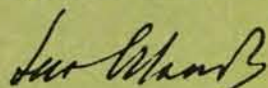
La mafia nella grande trasformazione

Tra la mafia tradizionale e la mafia di oggi c'è di mezzo la «grande trasformazione» postbellica, e cioè un imponente sovvertimento strutturale che ha mandato in pezzi - nel Nord come nel Sud dell'Italia - un intero ordine economico e sociale sostituendolo con un altro. I due motori della grande trasformazione sono stati, nel Mezzogiorno, l'emigrazione e l'intervento pubblico. Entrambi hanno determinato una crisi profonda del potere e del comportamento mafioso in Calabria e in Sicilia. L'emigrazione degli anni cinquanta e sessanta ha un effetto dirompente sulla struttura dei gruppi mafiosi, in quanto ne assottiglia le «basi umane» e ne interrompe il ricambio generazionale. Anche l'espansione degli impieghi pubblici nel Mezzogiorno apre dei vuoti nelle file dei giovani uomini d'onore e fa diminuire la competizione per il controllo delle risorse locali. La grande trasformazione dà luogo inoltre a un processo di cambiamento culturale generale che nelle aree mafiose del Mezzogiorno si esprime sotto forma di un fenomeno di «identificazione tra onore e ricchezza». Il possesso della ricchezza, che nella mafia tradizionale era considerato come una delle prove e come una delle conseguenze della capacità di farsi rispettare, nell'opinione popolare degli anni del dopoguerra diventa esso stesso un atto meritorio. La ricchezza diventa essa stessa onorevole e conferisce onore a chi la possiede, indipendentemente dalle «gesta» violente e dall'onore conseguito per mezzo di esse.

Il mantenimento della supremazia impone adesso la disponibilità di ricchezze e consumi crescenti. Assistiamo così a una rapida caduta della conflittualità su base onorifica e familistica, cui si accompagna un declino delle più importanti funzioni dell'autorità mafiosa. Ma la rottura del meccanismo mafioso dell'integrazione provocata dalla rivoluzione postbellica e la conquista di una nuova identità da parte degli uomini d'onore non sono stati fenomeni indolori. Essi hanno comportato un'operazione di svincolamento da tutte quelle limitazioni imposte ai mafiosi dal loro ruolo di mediatori, protettori, giudici. La trasformazione postbellica del fenomeno mafioso coincide con un processo di perdita quasi completa della «legittimità» del potere mafioso medesimo.

Due circostanze influenzano marcatamente la separazione dei mafiosi dalla propria tradizionale base di legittimazione. Da un lato accade che nello stesso momento in cui potere e successo economico crescono di importanza come mete collettive, lo sviluppo dei moderni partiti e movimenti di massa nel Mezzogiorno - e la dura opposizione da essi messa in atto nei confronti dei mafiosi - diminuiscono le possibilità di perseguire queste mete da parte degli ex-uomini d'onore che si vedono sottratti consensi, uomini e spazi di azione. Dall'altro, si verifica negli anni cinquanta e sessanta un processo di avocazione da parte dello stato del controllo della violenza fisica sul territorio nazionale che comporta il ritiro della quasi secolare delega alle élite locali meridionali per la gestione dell'ordine pubblico. Una serie di campagne repressive, che iniziano negli anni quaranta e si concludono venti anni dopo, e l'istituzione nel 1956 di una nuova legge per il confino di polizia significano uno sconvolgimento del potere mafioso più lungo temporalmente, e più grave nei suoi effetti immediati e di medio periodo, di quello verificatosi in Sicilia tra il 1926 e il 1936 con il prefetto Mori. La costituzione di un effettivo monopolio centrale della forza da parte dello stato culmina nel 1962 con la nascita di una commissione parlamentare di inchiesta sul fenomeno della mafia in Sicilia. L'attività di questa Commissione contribuì notevolmente alla trasformazione dei rapporti tra il potere politico e il potere mafioso. Termina negli anni sessanta l'ufficialità e la legalizzazione di quest'ultimo, che si trova così a perdere un ulteriore elemento di sostegno e di giustificazione generali.

L'effetto di questo insieme di fattori culturali, politici ed economici è che i mafiosi vengono sospinti in una posizione di devianza e di marginalità. Privi di delega ufficiale per la mediazione del conflitto e per la repressione del comportamento non conforme, non più blanditi e lodati

I capimafia tradizionali erano perfettamente consapevoli dell'importanza del proprio ruolo. Calogero Vizzini individuò il problema dichiarando in una celebre intervista: «Il fatto è che in ogni società ci deve essere una categoria di persone che aggiustano le situazioni, quando si fanno complicate». I casi di «complicazione delle situazioni» erano, in pratica, tutti gli episodi di non-rispetto di alcune norme tradizionali di comportamento, e tutti i numerosi casi di conflittualità attuale o potenziale tra individui e gruppi generati nella sfera delle relazioni socioeconomiche. Le parti interessate evitavano così la trasformazione di un conflitto di interessi in una contesa giudiziaria o in un conflitto d'onore, risparmiando tempo e risorse umane e materiali. In questi casi il mafioso agiva come una specie di giudice conciliatore, garantendo attraverso il suo intervento la ricostituzione dell'ordine violato, e cioè la conservazione del sistema di valori e di comportamenti dominanti. Questa fotografia di Calogero Vizzini è di Nicola Scafidi.

ASSEMBLEA COSTITUENTE

Roma,lì II Luglio 1948

Caro Coppola,

mi è pervenuto il fusto del Suo vino eccellente. Al ringraziamento orale aggiungo quello scritto e conto di portare meco lo squisito liquore nella mia villa di Campiglioni per bere alla Sua salute.

Mi creda cordialmente

Sig. FRANCESCO COPPOLA
PARTINICO (Palermo)

Nella situazione tradizionale, il rapporto tra il capomafia e l'uomo politico consiste di una relazione clientelare di tipo classico, nella quale l'elemento dello scambio di favori reciproci resta collegato a un codice di amicizia e di cavalleria. L'aspetto contrattuale della relazione viene a trovarsi come «immerso» in una rete di gesti e di comportamenti ostentatamente disinteressati. Quando nel 1948 Vittorio Emanuele Orlando, allora deputato all'Assemblea Costituente della Repubblica, scrive al mafioso Francesco Coppola, capo-elettore di Partinico, per ringraziarlo del vino inviatogli, questa ideologia dell'amicizia risulta particolarmente evidente.

pubblicamente dalle autorità, additati nei comizi e nei tribunali come nemici dell'ordine e dello sviluppo, gli uomini d'onore conoscono negli anni cinquanta e ancor più negli anni sessanta una profonda crisi di collocazione e di identità. Il loro ruolo si avvicina sempre più a quello del semplice delinquente, del moderno gangster urbano privo di radici e di consensi popolari, qualche volta sostenuto, qualche volta tollerato e qualche volta represso dalle autorità ufficiali.

Gli anni settanta e l'impresa mafiosa

La grande trasformazione si conclude all'inizio degli anni settanta. Tale decennio si caratterizza per la rottura dell'equilibrio che aveva governato il decennio precedente. A partire dal 1972-1973 il flusso migratorio dal Sud al Nord si affievolisce fino a ridursi a valori minimi. In breve tempo il saldo migratorio del Mezzogiorno diventa addirittura positivo, e tale si mantiene fino ai nostri giorni. Parallelamente, l'intervento statale cambia natura: non ha più effetti di unificazione, ma di disgregazione dell'economia e della società meridionale. Data la natura clientelare e sperequata della distribuzione di risorse pubbliche avvenuta negli anni cinquanta e sessanta, il tentativo di concentrare e di «finalizzare» queste risorse che comincia a manifestarsi verso l'inizio degli anni settanta determina ondate periodiche di malcontento e di tensioni di carattere territoriale interne al Mezzogiorno.

Nella società italiana ricomincia a manifestarsi una serie di particolarismi che sembravano essere stati definitivamente cancellati dai processi di omogeneizzazione dello sviluppo precedente. Si scopre che l'Italia non possiede ancora né un sistema economico né un sistema territoriale realmente unificati. Si individuano molteplici stratificazioni all'interno del sistema delle imprese (imprese «assistite», «sommerse» ecc.), all'interno del mercato nazionale del lavoro - che risulta frantumato in una serie di sottomercati più o meno indipendenti - e all'interno della struttura politica e sociale.

È nell'ambito di questo mutato scenario generale, dominato da una tendenza verso la disintegrazione socioeconomica, che maturano le condizioni per la nascita della mafia imprendrice. È dalla coincidenza del ruolo del mafioso con il ruolo di deviante avvenuta nel primo ventennio postbellico che occorre partire per interpretare il mutamento del fenomeno mafioso. Per la prima volta nella loro storia, infatti, i mafiosi si sono trovati in una situazione che li ha obbligati ad agire e a pensare al di fuori della cultura tradizionale.

Un certo numero di essi non ha accettato le condizioni di subalternità e di emarginazione imposte dal nuovo corso economico e politico e ha tentato di sviluppare una reazione di tipo innovativo. Una parte dei mafiosi siciliani e calabresi ha reagito alla perdita della sua posizione di supremazia concentrando tutte le sue energie nell'accumulazione del capitale. Non deve sfuggire lo stretto rapporto che si è instaurato ancora una volta tra una condizione di devianza e la nascita dell'orientamento imprenditoriale. È questo un *locus classicus* della sociologia di W. Sombart e T. Veblen: come per gli ebrei, gli eretici e gli stranieri durante il primo capitalismo, è stata la presenza di una situazione di marginalità che ha spinto i mafiosi verso un agire spregiudicato in vista del massimo profitto. In quanto «esclusi», il significato della ricchezza è diventato per i mafiosi più importante che per gli altri gruppi della popolazione, poiché il denaro e l'accumulazione hanno costituito per loro l'unica via per la riconquista del potere e dell'onorabilità.

La disintegrazione economica e istituzionale dell'Italia degli anni settanta ha agito da catalizzatore del processo di conversione dei mafiosi alla religione dell'accumulazione. La crisi del controllo statale della violenza - controllo che si era stabilito, come abbiamo visto, nel ventennio precedente e che, se aveva contribuito a trasformare il mafioso in deviante, lo aveva nello stesso tempo «costretto» in tale ruolo - ha svolto una funzione cruciale nella moltiplicazione delle possibilità di azione dei mafiosi. Essa ha permesso loro di riadattare a scopi di accumulazione tutto quel bagaglio di gesta sopraffattorie e violente che caratterizzavano il curriculum dell'uomo d'onore di trent'anni prima. Il massiccio ingresso dei mafiosi nella competizione industriale rappresenta un'innovazione radicale. Si tratta però di una innovazione che avviene sotto il segno della riemersione di molte delle caratteristiche più arcaiche di questi personaggi.

I più grandi capitali necessari per l'acquisto di merci, macchine, edifici e forza-lavoro vengono ottenuti accentuando antiche forme di conflitto: viene richiamato in vita il sequestro di persona, assieme a tutto il feroce bagaglio rituale a esso collegato. Il mafioso imprenditore è il prodotto di una stupefacente mutazione culturale che comporta il riutilizzo anche di molte antiche qualità individuali. Il gusto dell'impresa rischiosa, la mancanza di scrupoli, la capacità di prescindere dalle conseguenze immediate delle proprie azioni, sono tutti elementi tipici del vecchio uomo di rispetto che ritroviamo, opportunamente modificati, in questo nuovo personaggio.

Dall'incontro della mafia con l'imprenditorialità si è venuta a costituire un'istituzione originale: l'«impresa mafiosa». A essa deve essere attribuita gran parte della micidiale capacità espansiva dimostrata dalla mafia nell'ultimo decennio. Rispetto alla normale impresa concorrenziale descritta nei manuali di economia, l'impresa mafiosa gode di una indiscussa superiorità economica garantita - a parità di tutte le altre condizioni - da tre vantaggi competitivi fondamentali. Il primo di questi consiste nella creazione di un ombrello protezionistico intorno al mercato di pertinenza dell'impresa mafiosa tramite lo scoraggiamento della concorrenza. L'impresa mafiosa riesce ad assicurarsi appalti e merci a prezzi di favore senza essere esposta alla stessa pressione concorrenziale di cui devono tenere conto le altre imprese. La capacità di intimidazione del metodo mafioso - che va dalla semplice minaccia fino all'assassinio dei concorrenti - è tale da agire come una vera e propria barriera doganale. Il secondo vantaggio competitivo dell'impresa mafiosa consiste nella compressione salariale e nella maggiore fluidità della manodopera in essa occupata. La compressione salariale assume prevalentemente la forma dell'evasione dei contributi previdenziali e assicurativi e del non pagamento degli straordinari. L'organizzazione autoritaria del lavoro, tipica dell'azienda mafiosa - il cui personale di sorveglianza è spesso costituito da pregiudicati, elementi della delinquenza comune, diffidati dalla pubblica sicurezza e parenti del proprietario - scoraggia attivamente qualunque protesta degli operai. Il terzo vantaggio competitivo consiste nella maggiore solidità ed elasticità di tipo finanziario dell'impresa mafiosa rispetto alla piccola e media azienda non mafiosa. Il denaro necessario per gli investimenti del capitalista mafioso non proviene tanto dall'accumulo degli ordinari profitti aziendali, ma viene ricavato da tutta una serie di fonti straordinarie di guadagno finanziario. I rilevanti capitali generati nel circuito delle attività illegali del mafioso, infatti, vengono continuamente travasati nel circuito delle operazioni imprenditoriali legali.

La crescita della mafia imprenditrice ha implicato un cambiamento dei rapporti tra le grandi imprese esterne al sistema economico locale da una parte e i gruppi mafiosi siciliani e calabresi dall'altro. Mentre negli anni cinquanta e sessanta vigeva una situazione di conflitto tra il parassitismo mafioso, che si esprimeva nella richiesta di tangenti, guardianie ecc., e la razionalità aziendale che considerava tali pretese come un fattore di aggravio dei costi di produzione, negli anni settanta maturano le condizioni per una vera e propria alleanza. Le grandi imprese che si aggiudicano i lavori pubblici nel Sud trovano adesso conveniente in molti casi l'incoraggiamento dell'imprenditorialità mafiosa. I tradizionali motivi di tranquillità di esercizio si accompagnano adesso a ragioni di politica aziendale a vasto raggio. Le imprese mafiose sono quasi sempre espressione di gruppi di potere locale con i quali è utile allearsi, costituendosi come gruppo di pressione più ampio nei confronti dello stato, allo scopo di far salire artificialmente i costi degli investimenti pubblici nel Sud.

Il tipo del mafioso imprenditore e i limiti del capitalismo mafioso

Il «tipo» del moderno mafioso imprenditore presenta caratteristiche differenti, e per molti versi opposte, a quelle del «tipo» del mafioso tradizionale. L'origine sociale composita e il possesso di un discreto livello di istruzione sono due caratteristiche salienti dell'odierno mafioso. Questi non proviene più quasi esclusivamente dalle classi subalterne, ma anche e sempre più frequentemente dai diversi strati del ceto medio professionale e impiegatizio, oltre che dal commercio e dall'industria medesimi. La rapidità della carriera dell'imprenditore mafioso rappresenta un altro elemento di differenziazione nei confronti del mafioso tradizionale. Questi doveva attendere di solito molti anni per raggiungere il potere e l'influenza di un capocosca. L'età media dei principali mafiosi non supera oggi i 35-40 anni.

Una terza cruciale differenza tra il tipo dell'odierno mafioso e quello del mafioso tradizionale consiste nella diversa qualità dei suoi rapporti con gli altri mafiosi da un lato e con il sistema giuridico dall'altro. Non esiste più, oggigiorno, la distinzione tra la fase anomica - contrassegnata da frequenti scontri con i propri simili e con l'autorità statale - e la successiva fase legale della carriera mafiosa, caratterizzata dal riconoscimento formale e di fatto della propria posizione di comando. Il continuo scontro con il sistema giuridico e con i concorrenti sembra contrassegnare la carriera del mafioso imprenditore. La distaccata sicurezza del mafioso classico nei confronti dell'esito giudiziario delle vicende criminali di cui era protagonista ha ceduto il posto a una attenzione vigile e preoccupata. Se Vito Cascio Ferro poteva essere accusato 69 volte e 69 volte assolto grazie alla capacità - tipica di ogni capomafia classico - di influenzare la magistratura e di manipolare tutti gli stadi del processo penale, non altrettanto si può dire del mafioso imprenditore. L'odierna maggiore disomogeneità sociale e culturale dei magistrati italiani e l'aumentata complessità del sistema della giustizia penale hanno reso meno manovrabile da parte del potere politico o del potere mafioso l'intero *iter* giudiziario di un determinato caso. Una protezione efficace dalle leggi penali richiede di conseguenza un

CAMERA DEI DEPUTATI

Roma, lì 13/4/1951

Carissimo Don Ciccio-

L'ultima volta che ci vedemmo all'Hotel de Palme Lei mi diceva giustamente che a Partinico occorreva un Deputato Regionale giovane svelto ed amico ed a portata di mano degli amici. L'amico Totò Motisi risponde a tutti questi requisiti ed io ho deciso di aiutarlo con tutte le mie forze. Se a Partinico mi aiutate lo faremo diventare Deputato.

Con affettuosi saluti
mi creda

(Giovanni Palazzolo)

Sig.Francesco Coppola
Partinico

In cambio del loro appoggio elettorale, i mafiosi ricevevano dagli uomini politici favori per se stessi e per la propria cosca. L'integrazione per via mafiosa delle subculture politiche locali nel sistema politico nazionale si è costantemente ingrandita a partire dagli anni ottanta del secolo passato, con l'avvento al potere della Sinistra storica e con la prima importante estensione del suffragio elettorale (1882). Dopo il 1912, con l'introduzione del suffragio universale maschile, l'appoggio dei gruppi mafiosi nel corso delle campagne elettorali divenne una delle condizioni di maggior importanza per la conquista dei seggi parlamentari da parte dei candidati appartenenti ai partiti di governo.

considerevole investimento di risorse monetarie e un bagaglio non indifferente di conoscenze tecniche. Da qui la tendenza di ciascun gruppo mafioso di rilievo a stabilire rapporti di consulenza e di collaborazione fissa con avvocati, esperti fiscali e operatori giudiziari in grado di fornire informazioni e consigli adeguati. La figura del *consigliori,* dell'avvocato che è membro interno di una cosca mafiosa, resa popolare dal libro *Il Padrino,* è sempre più diffusa in Calabria e in Sicilia.

Per tutta una fase dello sviluppo della mafia imprenditrice, lo stretto legame del mafioso con il territorio e la cultura locale ha costituito un formidabile strumento di potere economico. Questi stessi legami, però, oltre un certo livello diventano delle catene per la libera espansione della sua imprenditorialità. Con lo sviluppo dell'accumulazione mafiosa, il monopolio territoriale di una data attività economica diventa un vestito sempre più stretto. La riproduzione allargata del proprio capitale attraverso investimenti che accrescono molto rapidamente la scala precedente di attività fa sì che l'invasione delle zone di altrui pertinenza diventi una necessità permanente dell'imprenditore mafioso. Il conflitto continuo diventa perciò inevitabile e l'accumulazione mafiosa si viene a scontrare con le basi territoriali del potere mafioso stesso.

Nella «società di transizione permanente», rinvenibile in forma concentrata nella Piana di Gioia Tauro *(zona in colore)*, in Calabria, e in alcune aree della Sicilia occidentale prima degli anni cinquanta, il comportamento mafioso costituiva la forma fondamentale dell'azione sociale e il potere mafioso giocava un ruolo decisivo nel processo dell'integrazione socioeconomica.

Non bisogna inoltre trascurare un altro fatto. La superiorità concorrenziale dell'impresa mafiosa, assicurata dai tre vantaggi competitivi che abbiamo descritto, viene ad annullarsi nei confronti di un'altra impresa di tipo mafioso, la quale può godere degli stessi vantaggi. Il risultato è uno scontro il cui criterio di regolazione non è l'efficienza produttiva, ma quella coercitiva. L'imprenditorialità mafiosa non viene però limitata solo dallo squilibrio tra la rapidità dell'accumulazione e l'insufficiente libertà dei mercati territoriali. Viene anche limitata dalla stessa particolare cultura dei mafiosi, che tende a trasformare il confronto impersonale di mercato in confronto antagonistico tra persone. La combinazione tra tradizionalismo e imprenditorialità, infatti, si sta rivelando come una miscela esplosiva. Un conflitto economico tra due imprese - che in altri contesti non provoca alcuna conseguenza al di fuori della sfera produttiva o distributiva - diventa qui un conflitto tra due «comunità politiche» sovrane che si trasforma subito in una guerra tra clan e famiglie che coinvolge decine e anche centinaia di parenti, amici e clienti. L'ascesa della mafia imprenditrice ha finito con il realizzare l'autentica guerra di tutti contro tutti, con morti, feriti e dispersi. Il bilancio di questa guerra è di quasi 1000 morti in Calabria e in Sicilia dal 1970 a oggi. Lungi dal determinare una progressiva estensione della sfera dei valori e delle condotte di tipo razionale-capitalistico, la pratica imprenditoriale dei mafiosi ha l'effetto di accrescere l'ampiezza della sfera delle attitudini arcaiche e predatorie presenti nella loro cultura e nella loro storia.

Mafia e mercato della droga

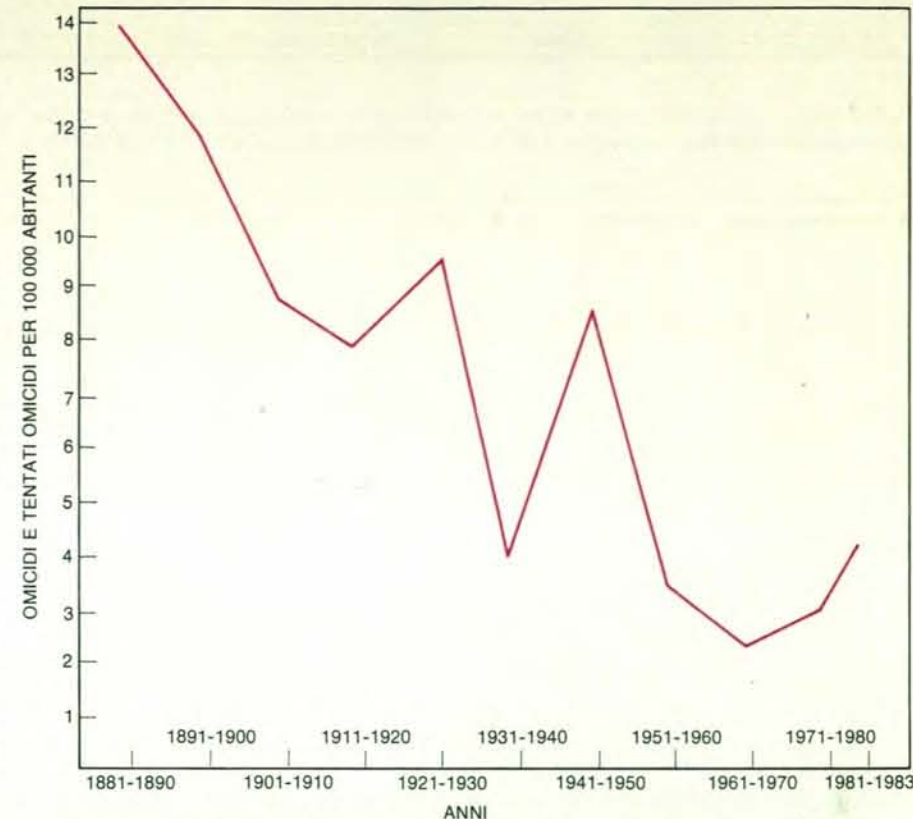
Il grande potere di mercato detenuto dagli odierni gruppi mafiosi della Sicilia occidentale sulla via dell'eroina costituisce un fatto nuovo, legato anch'esso ai mutamenti avvenuti negli ultimi quindici anni. Fino all'inizio degli anni settanta la conquista di una posizione oligopolistica sul mercato della droga era impedita ai mafiosi siciliani dalla loro insufficiente competitività nei confronti degli altri gruppi criminali europei, e in particolare di quelli francesi aventi base a Marsiglia.

I gruppi criminali che operano sul mercato dell'eroina possono essere considerati delle imprese, il cui successo di mercato dipende dalla possibilità di disporre e di organizzare tre beni e servizi fondamentali: capitale, violenza e non-azione della polizia e della magistratura. Esiste uno stretto rapporto tra la disponibilità di questi elementi, da un lato, e il prezzo e la quantità di droga esportata da ciascuna impresa criminale, dall'altro. Disponibilità di capitale significa acquisto di grandi partite di morfina base da trasformare in eroina tramite laboratori chimici clandestini gestiti dagli stessi gruppi mafiosi. Uso della violenza significa in primo luogo possibilità di costituire un ombrello protezionistico intorno al mercato di pertinenza dell'impresa criminale, in termini di barriere di ingresso che scoraggino eventuali concorrenti. In secondo luogo, significa disporre di personale specializzato (*killer* ecc.) in grado di eliminare ostacoli imprevisti: testimoni scomodi, membri del gruppo criminale che diventino informatori della polizia, magistrati e poliziotti che deviino dallo standard investigativo vigente in loco.

Non-azione della polizia e della magistratura vuol dire instaurazione di uno standard molto elevato di segretezza delle operazioni di compravendita, di trasformazione e di trasporto. Il sequestro di un quantitativo particolarmente alto di droga o di capitale, l'arresto di uno o più membri-chiave della stessa famiglia e, soprattutto, la scoperta dei laboratori clandestini costituiscono degli ostacoli alla normale conduzione degli affari che possono determinare la estromissione dal mercato di una impresa criminale e la sua sostituzione con un'altra. La non-azione dell'autorità inquirente può essere ottenuta tramite corruzione e pressioni politiche, oppure - come nel caso di Palermo negli ultimi anni - si può tentare di ottenerla tramite l'effetto terroristico di omicidi di alte autorità politiche e giudiziarie impegnate in attività antimafia.

La necessità di una presenza contemporanea di questi tre elementi fa sì che il numero ottimale delle famiglie-imprese criminali impegnate negli stadi più alti del *business* della droga si restringa fino a ridursi a poche unità per ciascuna grande corrente commerciale.

Ai fini del nostro discorso importa rilevare come lungo tutto l'arco degli anni cinquanta e sessanta la posizione delle famiglie-imprese siciliane restò contrassegnata da un costante svantaggio rispetto ai gruppi concorrenti soprattutto sotto l'aspetto della disponibilità di capitale. Prima dell'avvento della mafia imprenditrice, le funzioni di mediazione e la «cultura dell'onore», tipiche della maggioranza dei mafiosi, precludevano l'accumulazione di quelle grandi somme di denaro contante necessario per un ingresso in grande stile nel circuito dell'*import-export* di eroina su scala mondiale.

L'emergere - in Sicilia verso l'inizio degli anni settanta - di una schiera di mafiosi imprenditori e di piccoli e grandi speculatori non-mafiosi caratterizzati da un forte eccesso di liquidità e dalla ricerca di nuovi sbocchi di investimento, ha capovolto la posizione di svantaggio dei mafiosi siciliani sul mercato internazionale della droga.

Se adottiamo il numero annuo degli omicidi che avvengono in una data società come l'indicatore più sintetico del grado di controllo statale della violenza in essa vigente, la situazione dell'Italia degli anni settanta e dell'inizio degli anni ottanta emerge con caratteri propri di eccezionalità rispetto non solo al ventennio precedente, ma a tutto un lungo periodo della nostra vita nazionale. Il *boom* della conflittualità omicida che contrassegna gli ultimi quattordici anni testimonia l'interruzione di una tendenza quasi secolare di diminuzione degli omicidi. Secondo il *Sommario di statistiche storiche* pubblicato dall'ISTAT, il tasso degli omicidi e dei tentati omicidi si è ridotto in Italia di ben cinque volte tra il decennio 1881-1890 e il decennio 1961-1970, passando da un valore di 13,9 a uno di 2,6 per ogni 100 000 abitanti. La diminuzione del numero degli omicidi è stata tumultuosa e ha conosciuto solo due interruzioni prima degli anni settanta. Tali interruzioni sono avvenute entrambe in concomitanza delle due guerre mondiali. Dal 1970 al 1975 le denunce di omicidio si sono accresciute in Italia del 28 per cento. Tra il 1975 e il 1977 sono aumentate ancora del 27 per cento, superando quota 2000 e continuando a crescere fino a raggiungere la cifra di 2400-2500 negli anni a noi più vicini. Il grafico è stato elaborato su dati ISTAT.

L'accresciuta quantità di capitale di marca mafiosa, sommandosi alla parallela crisi del controllo statale della violenza e al conseguente incremento della non-azione della polizia, ha fatto sì che i mafiosi siciliani conquistassero in breve tempo un potere di mercato talmente esteso da modificare verso la metà degli anni settanta una divisione internazionale del lavoro criminale che vigeva da oltre un trentennio, spostando il baricentro della rotta dell'eroina dalla Francia all'Italia.

Le cinque famiglie criminali più potenti sono riuscite così a impiantare altrettanti laboratori per la produzione di eroina nei dintorni di Palermo, dotati ciascuno di una potenzialità produttiva di 50 chilogrammi alla settimana. Tenuto conto della irregolarità dei rifornimenti di morfina e dei problemi di clandestinità dell'industria, si può stimare - per il periodo che va dal 1975-1976 alla scoperta dei laboratori nell'agosto 1980 - una produzione annua di 4-5 tonnellate di eroina pura. Queste rappresentano circa il 30 per cento dell'intero fabbisogno americano, e venivano distribuite per oltre l'80 per cento nella città di New York, tramite una rete di corrieri, importatori e grossisti che coinvolgeva diverse centinaia di persone. Detratti i costi di produzione e di trasporto, la cifra di utile netto annuo si aggirava sui 1000 miliardi di lire. Si tratta di una grande quantità di denaro, concentrata in poche mani e dotata di alcune caratteristiche particolari. Le narco-lire, infatti, provengono dall'esterno del sistema economico locale e costiuiscono, quindi, una voce attiva della bilancia siciliana dei pagamenti; non sono di origine statale e vengono conservate e distribuite secondo canali indipendenti dal clientelismo politico classico. Esse hanno agito come un ulteriore fattore di decomposizione della società e delle istituzioni da un lato, e dall'altro lato come fattore di creazione di un nuovo consenso attorno ai mafiosi e ai gruppi e ai partiti legati alla mafia.

L'autonomia politica del potere mafioso

La crescita della mafia imprenditrice e la conquista di una posizione oligopolistica da parte delle famiglie-imprese mafio-

COSCHE DELLA CALABRIA TIRRENICA		COSCHE DELLA CALABRIA JONICA	
NOME DELLA COSCA	NUMERO DI FAMIGLIE DI FRATELLI	NOME DELLA COSCA	NUMERO DI FAMIGLIE DI FRATELLI
DE STEFANO	3	NIRTA	3
MAMMOLITI	4	MAZZAFERRO	6
RUGOLO	6	CATALDO	4
CIANCI	3	D'AGOSTINO	4
PIROMALLI	3	URSINO	6
MAZZAFERRO	4	JERINÒ	4
PESCE	4	RUGA	4

A differenza della mafia tradizionale, nella quale il potere del capo dipendeva anche dalla sua capacità di creare attorno alla propria persona un seguito più o meno permanente di soggetti a lui legati da una gamma relativamente ampia di rapporti, e che non oltrepassava quasi mai le 15-20 unità, l'odierna mafia imprenditrice consiste di una serie di vasti raggruppamenti di persone - che possono arrivare a includere oltre 70-80 elementi maschili adulti - con al centro un nucleo formato da una o più famiglie biologiche di dimensioni insolitamente grandi. La parentela biologica di primo grado tende con il tempo a sostituire gli altri tipi di relazione e a divenire l'esclusivo legame interno dei moderni gruppi mafiosi. Un'apposita indagine effettuata su 14 tra quelle che risultavano essere le più potenti cosche mafiose calabresi nel 1979-1980 ha portato a un risultato sorprendente: nessuna di esse risultava composta, nel suo nucleo fondamentale, da meno di tre fratelli. Ragioni di carattere economico legate alla mutata natura dell'attività mafiosa - basata oggi su famiglie che sono anche imprese e su imprese che sono anche famiglie - contribuiscono a spiegare il fenomeno in oggetto. Esistono tuttavia anche importanti ragioni di ordine «militare», connesse all'aumentata frequenza dei conflitti interni ai gruppi mafiosi e degli scontri con l'autorità giudiziaria, che spingono verso l'aumento del numero dei legami di parentela biologica necessari sia per il mantenimento sia per la conquista di una posizione di preminenza nell'ambito del microcosmo mafioso.

se nel sistema mondiale dell'eroina hanno avuto l'effetto di accrescere l'autonomia e il peso del potere mafioso rispetto al potere politico. L'aumentata autonomia di azione dei gruppi mafiosi non va però considerata come il prodotto passivo dei cambiamenti economici. Essa ha interagito con questi ultimi, contribuendo a crearli e a rafforzarli non meno di quanto ne sia stata determinata.

Il fenomeno dell'autonomia politica del potere mafioso costituisce uno dei tratti salienti della situazione attuale e rappresenta un elemento di novità rispetto ai precedenti modelli di rapporto mafia-politica. È avvenuta una rottura del rapporto di subordinazione vigente tra il capomafia locale e l'esponente politico governativo. Il rapporto con il mondo esterno alla zona di influenza del mafioso non è più mediato dal deputato locale o dal locale ministro o sottosegretario, ma è diretto e immediato. I mafiosi imprenditori hanno allargato il proprio potere entrando in rapporto diretto non solo con il personale politico e amministrativo regionale e meridionale, ma con tutto un settore paralegale e illegale dell'apparato centrale dello stato.

Due sono le forme principali attraverso cui si manifesta l'autonomia del potere mafioso: l'«internalizzazione» della rappresentanza politica nei gruppi mafiosi; la costituzione di rapporti di comune interesse economico tra leader mafiosi, leader politici e settori del mondo economico e finanziario locale e nazionale tramite lo strumento della «*lobby* politico-mafiosa».

Il fenomeno della internalizzazione della rappresentanza politica consiste nella tendenza, tipica dei principali gruppi mafiosi, a mobilitare il proprio peso elettorale in favore di propri membri interni che si presentano come candidati, o ad appoggiare persone legate da stretti vincoli di parentela al capo o ai capi della cosca. Tale soluzione presenta il vantaggio di semplificare il problema dei rapporti con il potere ufficiale, eliminando la necessità di dover ricorrere a trattative, pressioni, accordi e alleanze con uomini politici esterni alla famiglia mafiosa medesima. Gli svantaggi consistono in una esposizione pubblica troppo evidente e in una troppo accentuata concentrazione di potere che possono divenire controproducenti nei momenti critici, di scontro frontale con l'opinione pubblica e con le istituzioni giudiziarie.

L'ampio disinteresse dell'opinione pubblica nazionale e dei poteri centrali nei confronti del problema della mafia durato dall'inizio degli anni settanta alla fine del 1982, con l'assassinio a Palermo del generale Dalla Chiesa, ha favorito, però, un vasto processo di identificazione di ruoli politici e di ruoli mafiosi che, partito dal basso, è arrivato progressivamente a condizionare la vita delle amministrazioni provinciali e regionali calabresi e siciliane.

Una lobby politico-mafiosa consiste di un gruppo di persone inserite in una serie di posizioni-chiave della vita economica, politica e istituzionale unite da una trama di obbligazioni reciproche allo scopo di monopolizzare le risorse e i servizi strategici di una data comunità o di un dato settore. Tale istituzione è tipica dell'ambiente mafioso e non deve essere confusa con forme apparentemente simili di aggregazione quali i gruppi clientelari e di pressione. A differenza dei rapporti vigenti all'interno dei gruppi clientelari, le relazioni tra i membri delle lobby politico-mafiose sono caratterizzate da un ampio egualitarismo. Si tratta di gruppi privi di una gerarchia più o meno formale e non centralizzati.

La lobby politico-mafiosa si differenzia dal gruppo di pressione per la sua composizione eterogenea, tendente ad aggirare i limiti al potere di ciascuno imposti dalla divisione dei poteri. Imprenditori, uomini politici, mafiosi, autorità giudiziarie, dirigenti di enti pubblici, trovano nei rapporti di scambio vigenti all'interno di queste istituzioni un modo per convertire il contenuto specifico del proprio potere in altro contenuto di altro potere.

Uno dei veicoli più importanti della progressiva penetrazione del potere mafioso nei gangli più delicati delle istituzioni pubbliche e della vita economica nazionale è consistito nella alleanza - di cui abbiamo già parlato - tra la grande impresa legata agli appalti pubblici e la mafia. Un secondo strumento dell'integrazione del potere mafioso nell'*élite* del potere nazionale è costituito dalla alleanza tra i gruppi mafiosi finanziariamente più forti e un'intera sezione del capitale finanziario nazionale e internazionale. Tutto il capolavoro dell'avvocato Sindona è consistito appunto nella creazione di un canale di comunicazione - inesistente prima degli anni settanta - tra il circuito finanziario legale ed il circuito dei capitali accumulati illegalmente.

Piccole grandi lobby politico-mafiose si sono diffuse dappertutto nel nostro paese lungo l'arco dell'ultimo quindicennio, fornendo lo sfondo per la formazione di coalizioni e di cointeressenze economiche proibite dalla legge. Sono innumerevoli i casi di società immobiliari e finanziarie, compagnie di assicurazione, imprese commerciali, industriali e agricole, nonché i casi di investimenti nel settore turistico-alberghiero gestiti da noti mafiosi associati a noti uomini politici.

Legge anti-mafia e sviluppo economico del Mezzogiorno

L'analisi fin qui svolta ha fornito le coordinate teoriche della nuova legge antimafia dello stato italiano, approvata nel settembre 1982, a pochi giorni dall'assassinio del generale Dalla Chiesa, e detta «legge La Torre» dal nome del suo presentatore, il deputato comunista Pio La Torre, assassinato anch'egli dalla mafia alcuni mesi prima di Dalla Chiesa. Si tratta di una legge molto dura - la cui applicazione è ancora largamente disattesa dalle autorità giudiziarie dell'Italia meridionale - il cui punto focale consiste nella possibilità di confiscare i patrimoni mafiosi di cui non è stata dimostrata l'origine lecita e nell'allargamento delle possibilità di indagine nelle banche e nelle istituzioni finanziarie.

Quali possono essere le conseguenze di una applicazione severa della legge La Torre sul futuro dei gruppi mafiosi e sullo sviluppo economico del Mezzogiorno?

È importante rilevare al proposito come

La dimostrazione più chiara della potenza di mercato dell'impresa mafiosa è stata fornita dalla famosa vicenda degli appalti relativi alla costruzione del porto di Gioia Tauro. I mafiosi imprenditori sono riusciti ad avere la meglio nella competizione per i subappalti, assicurandosi la gestione diretta del 70 per cento di questi e il controllo indiretto del resto tramite una tangente dell'8 per cento sull'importo di ciascun subappalto eseguito dalle imprese non mafiose. La ricostruzione di tutti i pagamenti ufficiali (eseguiti dai due consorzi di grandi imprese appaltanti) agli autotrasportatori effettuata dal giudice Cordova permette di individuare una caratteristica rilevante della dinamica della mafia imprenditrice: l'ordine e la stabilità interni a quest'ultima vengono raggiunti quando la gerarchia del potere economico viene a coincidere con la gerarchia del prestigio e del potere militare vigente tra le sue unità costitutive. La fotografia (dello Studio Victor) è pubblicata per concessione della DIPENTA S.p.A.

La vastità degli interessi in gioco ha trasformato quasi completamente l'atmosfera dei processi penali che si svolgono oggi nelle zone mafiose. Lo spazio per le dotte dissertazioni di diritto e per il cerimoniale cavalleresco tra accusa e difesa si è ridotto al minimo o si è chiuso. Un drammatico, duro scontro di posizioni costituisce la nota dominante dei moderni processi di mafia. L'oratoria degli avvocati è sobria e aggressiva. Gli imputati sono in grado di capire tutti i passaggi e di afferrare tutte le sfumature del dibattimento, suggerendo qualche volta le linee difensive, valutando il comportamento di ciascuno. I più stretti legami vigenti oggi tra avvocati e mafiosi affiorano a volte nel linguaggio dei gesti e dei segni che circolano nell'aula tra i banchi, o le «gabbie», degli imputati e le sedie della difesa. Ma affiorano anche nei modi di esprimersi degli avvocati stessi. Non è difficile ascoltare frasi come queste: «Signori della Corte! L'innocenza di questo imputato si impone con la stessa perentorietà di un colpo di Magnum 44!». La fotografia, di Giovanna Borgese, è stata ripresa al processo «Mafia e droga» nel novembre 1982 a Palermo.

uno degli elementi cruciali della espansione mafiosa nell'economia meridionale e italiana sia costituito dal livello molto alto di interrelazione vigente tra il settore legale e il settore illegale delle economie. È proprio l'investimento legale di capitali acquisiti illecitamente che rappresenta una delle differenze più importanti tra l'odierno fenomeno mafioso in Italia e il gangsterismo e la mafia americani. La maggior parte dei profitti conseguiti dagli imprenditori del crimine statunitensi tramite l'organizzazione del gioco d'azzardo, del traffico della droga, della prostituzione e del *racket*, viene riversata nello stesso mercato illegale. Le proporzioni di quest'ultimo sono sufficientemente ampie da assorbire investimenti anche di grandi proporzioni: si pensi al caso delle città di Las Vegas e di Atlantic City, al giro d'affari delle scommesse clandestine e ad altri fatti ancora.

La possibilità di travasare grandi quantità di capitale dalla sfera illegale a quella legale risulta inoltre drasticamente limitata negli Stati Uniti dall'esistenza di apposite barriere istituzionali e informali, costituite dalle leggi sui capitali di sospetta provenienza, dalla labilità del segreto bancario e dai controlli e dagli interventi anche «pesanti» della polizia e dell'FBI sulle attività economiche della criminalità organizzata.

Alla luce di quanto detto finora, è possibile intravvedere alcune delle conseguenze possibili della legge La Torre sulla dinamica dei due settori in cui operano le imprese mafiose italiane. La nuova legge può costituire quella «barriera» tra il mercato legale e quello illegale che è mancata finora nell'economia meridionale e nazionale. In presenza della legge, i due mercati continuerebbero a funzionare in condizioni radicalmente differenti. Verrebbe a esaurirsi o a ridursi al minimo quel serbatoio di risorse aggiuntive rappresentato dall'uso del capitale e dei servizi di origine mafiosa e criminale nell'economia legale, e verrebbero con ciò ristabilite più eque e omogenee condizioni della concorrenza tra imprese mafiose e imprese non mafiose sui mercati del lavoro, delle merci e della moneta.

La separazione tra i due settori coinciderebbe inoltre con la costituzione di una «gerarchia» tra di essi. La quota di rischio inerente a ciascuna operazione economica illegale aumenterà, e avverrà un deflusso di capitali e di investimenti dal settore illecito. Costringendo gli imprenditori mafiosi a dividere la propria attività, e in una certa misura anche la propria identità, tra l'economia del crimine e l'economia legale, tra il capitalismo mafioso e d'avventura e il capitalismo *tout court*, in condizioni complessive di tipo deflazionistico, la legge La Torre può avere l'effetto di innescare quella tendenza verso la legalizzazione che è finora mancata nella vicenda della mafia italiana. Non bisogna infatti dimenticare come gli imprenditori mafiosi - a differenza degli imprenditori e degli uomini d'affari «normali» - non possono contare su una sicurezza degli scambi e su diritti di proprietà garantiti da leggi e istituzioni formali, e ciò rappresenta di per sé una pesante limitazione all'espansione di qualunque attività economica di natura illegale.

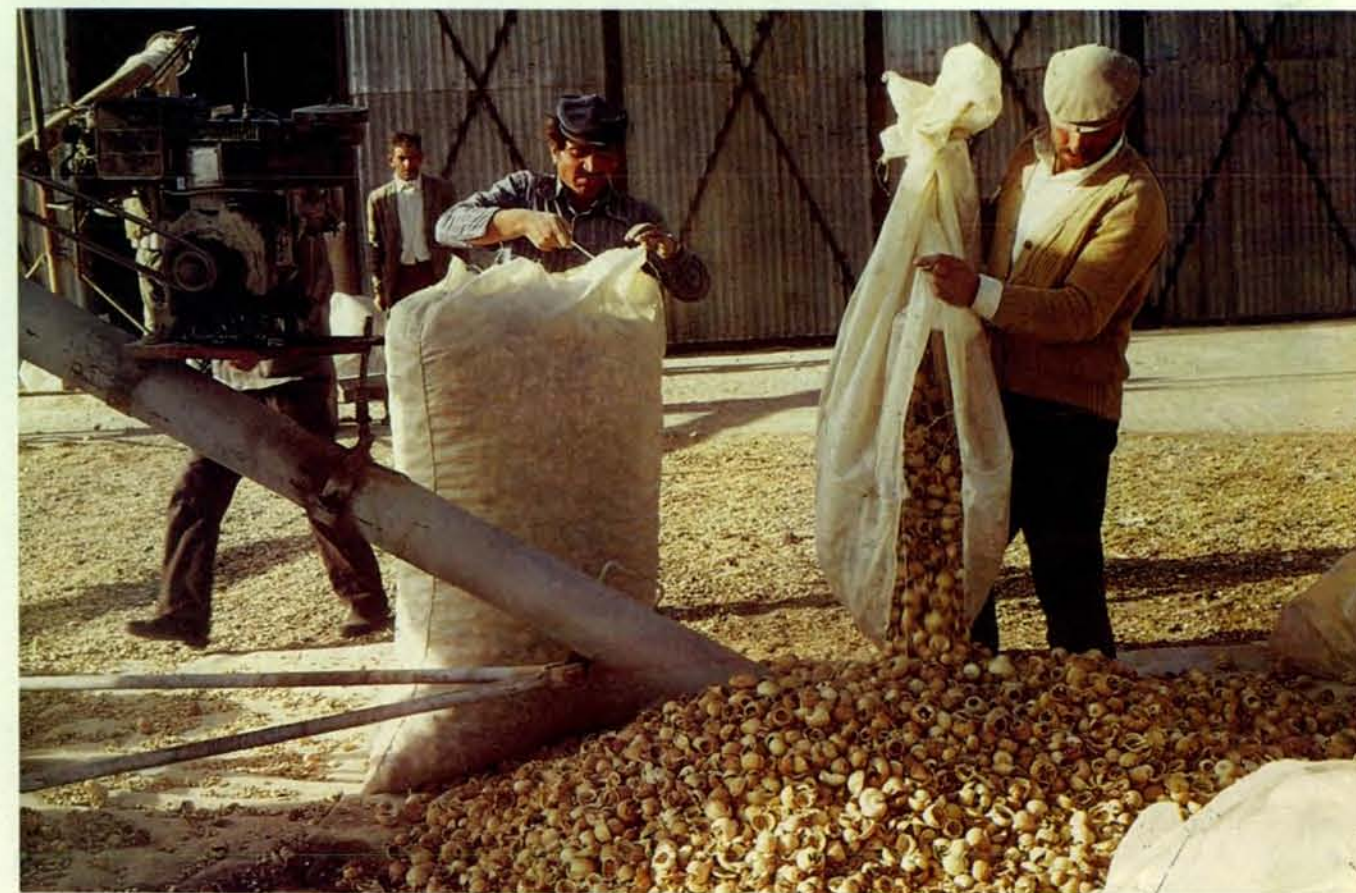
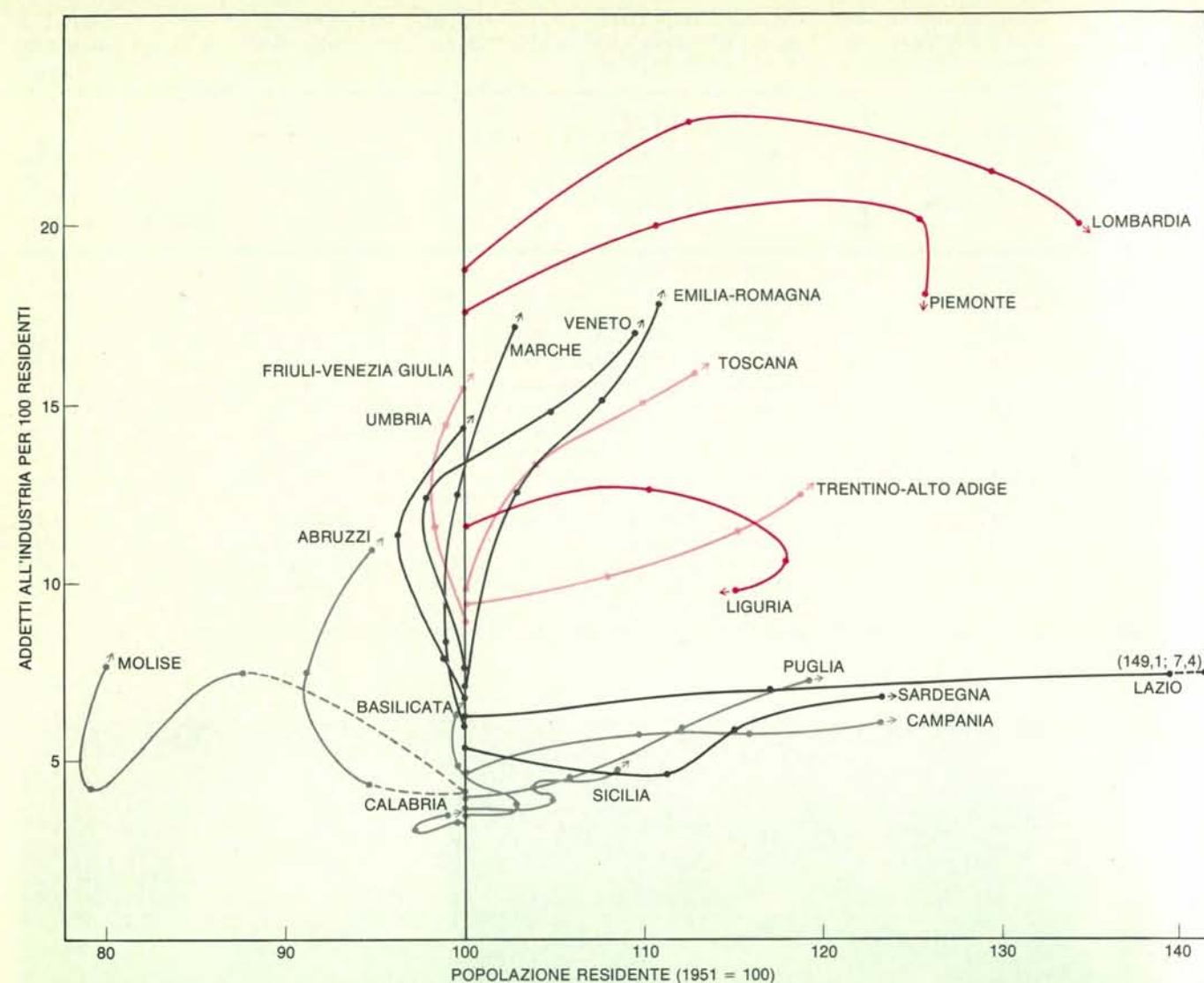
Non ha, infatti, molto senso continuare a reinvestire all'infinito nel mercato illecito. Anche se i tassi di remunerazione del capitale investito sono molto alti, il rischio inerente all'esercizio di una attività imprenditoriale che non è legalmente tutelata diventa a un certo punto così alto da innescare quasi naturalmente una tendenza verso la conversione dei patrimoni e delle attività illegali in titoli di proprietà e beni protetti dalle leggi ufficiali. La presenza di una efficace barriera tra i due circuiti economici in cui operano le imprese mafiose può avere quindi il duplice effetto di «tenere sotto controllo» il ritmo di formazione e il processo di riproduzione della ricchezza criminale da un lato, e di ridurre entro limiti definiti la quota dei beni di origine illegale che diventa legittima e che può essere negoziata e trasferita per eredità, dall'altro.

Nel caso di una applicazione rigida della legge La Torre, il tasso di rischio dell'investimento illegale diventerebbe talmente alto da imporre alle famiglie-imprese mafiose un rapido trasferimento di risorse dal settore criminale a quello legale, anche allo scopo di rendere trasmissibile ai propri discendenti la maggiore quantità possibile di queste stesse risorse.

La problematica dei possibili effetti della legge La Torre sulle *chance* di sviluppo economico del Mezzogiorno va affrontata, oltre che per la sua rilevanza intrinseca, anche per un confronto con una tesi recentemente diffusa da intellettuali e uomini politici siciliani, secondo la quale la messa in atto della nuova legge finirebbe per provocare - date le presenti condizioni di dominio mafioso di vasti rami dell'attività economica - una seria contrazione della produzione, degli investimenti e dell'occupazione nelle regioni meridionali caratterizzate dalla presenza della mafia e della camorra.

Questa tesi si appoggia, ad avviso di chi scrive, su una analisi sbagliata del fenomeno mafioso e trascura alcuni fatti economici tra i più evidenti. In primo luogo, essa non considera il fatto che la crescita delle imprese mafiose non è avvenuta «in concomitanza» a un parallelo sviluppo delle imprese preesistenti, ma è consistita in larga parte in un processo di sostituzione delle prime alle seconde, costrette a finanziare il capitalismo mafioso tramite il pagamento di tangenti, l'esclusione dai mercati di vendita e dagli appalti più redditizi e la rinuncia ai programmi di investimento nelle aree geografiche ed economiche coperte dall'ombrello protezionistico della mafia o della camorra. L'impresa mafiosa non ha attivato risorse produttive prima inutilizzate, ma ha combinato e sfruttato in modo diverso dei mezzi di produzione già esistenti.

In secondo luogo occorre sottolineare come la crescita del capitalismo mafioso non sia avvenuta nel corso di un vasto processo di sviluppo economico. Le aree meridionali in cui essa si è manifestata con particolare intensità non figurano tra quelle contrassegnate dai maggiori tassi di crescita della produzione e degli investimenti. È possibile semmai rinvenire un rapporto di proporzionalità inversa tra la presenza di fenomeni mafiosi, camorristici e di grande criminalità da un lato, e tassi di sviluppo delle economie locali dall'altro. Le due regioni meridionali che hanno sperimentato i più elevati indici di crescita dell'attività economica nel corso dell'ultimo quindicennio - e cioè l'Abruzzo e la Puglia - sono anche quelle che mostrano tassi di criminalità organizzata e comune tra i più bassi del Mezzogiorno. Le poche *enclave* di sviluppo agricolo e manifatturiero presenti in Calabria e in Sicilia sono collocate nelle zone non-mafiose (o «meno mafiose», se si preferisce) delle due regioni, e cioè nella provincia di Cosenza e nella Sicilia orientale.

L'ascesa della mafia imprenditrice ha ridotto le possibilità di sviluppo economico autoctono delle zone in cui si è manifestata. Non ha bloccato del tutto lo sviluppo delle forze produttive, ma ne ha deviato e rallentato il cammino. Non trattando-

Il mercato dell'eroina è caratterizzato da uno squilibrio permanente tra domanda e offerta a vantaggio della seconda, che gode di profitti stabili ed elevati. La struttura della domanda è quanto di più «ideale» si possa immaginare: una massa amorfa di consumatori privi di qualunque possibilità di influenzare il mercato e che tendono comunque ad acquistare la merce, essendo costretti a prescindere dalla sua qualità e dal suo prezzo. La struttura dell'offerta è più complessa e può essere divisa in due sezioni fondamentali: il settore «competitivo», nel quale opera una serie di unità indipendenti di piccole e medie dimensioni la cui attività consiste nel fornire eroina a un pubblico di non criminali; e un settore «oligopolistico», costituito da un numero limitato di famiglie-imprese criminali di diversa nazionalità, la cui attività consiste nella vendita di droga all'ingrosso ai membri della sezione precedente. La visibilità dell'intervento mafioso e della grande criminalità nell'economia dell'eroina è, quindi, inversamente proporzionale all'ampiezza dell'intervento medesimo. Nella fotografia (Sipa/Grazia Neri) una fase della raccolta delle capsule di *Papaver somniferum* da cui si estrae l'oppio; dall'oppio si ricavano successivamente morfina ed eroina.

Se prendiamo come indicatore del tasso di sviluppo il numero degli addetti all'industria per ogni 100 residenti ai censimenti 1951, 1961, 1971, 1981 *(pallini sulle curve)*, possiamo constatare come le sole tre regioni italiane rimaste al di sotto del 6 per cento - in condizioni globali, perciò, di stagnazione e di non-industrializzazione lungo l'arco del trentennio - sono proprio quelle del «triangolo criminale» del Mezzogiorno: Calabria, Sicilia e Campania. Nel periodo 1951-1981, nel corso di tre ondate successive di crescita industriale - come mostrato da G. Becattini e G. Bianchi nel loro originale studio *Sulla multiregionalità dello sviluppo economico italiano* - tutte le altre regioni si sono avviate, più o meno speditamente, sulla strada dello sviluppo industriale. Nella figura qui sopra si può infatti vedere come, allineando le regioni rispetto alla popolazione residente nel 1951, e collocandole su una scala basata sui differenti livelli di industrializzazione, è possibile identificare quattro gruppi di regioni: «In testa figura il gruppo delle regioni più industrializzate (oltre il 10 per cento di addetti all'industria): Lombardia, Piemonte e, distanziata, Liguria *(in colore intenso)*; segue, con valori attorno al 10 per cento, un terzetto di regioni, diciamo mediamente industrializzate: Toscana, Friuli-Venezia Giulia, Trentino-Alto Adige *(in colore chiaro)*; nello spazio tra il 6 e il 7 per cento si raggruppano le regioni, se così si può dire, scarsamente industrializzate: Veneto, Emilia-Romagna, Marche, Umbria, Sardegna, Lazio *(in nero)*; le restanti regioni si possono definire non-industrializzate... Si registrano, infine, le posizioni al (provvisorio) traguardo del 1981: le regioni della prima (Toscana, Emilia-Romagna, Veneto) e della seconda (Marche, Umbria) ondata postbellica esibiscono livelli di industrializzazione ormai paragonabili a quelli delle due "grandi" originarie, pur nelle evidenti differenziazioni...; le regioni della terza ondata, Abruzzi, Molise, Basilicata e Puglia (la sola in costante crescita di popolazione) hanno raggiunto e qualche volta superato quelli che erano i livelli di industrializzazione delle precedenti regioni al termine del primo periodo; Sardegna e Lazio seguono diversi, ma fra loro distinti, percorsi, pur accomunate da una espansione demografica, di intensità differenziata, ma senza soste nel trentennio considerato; dopo trent'anni, Campania, Calabria e Sicilia sembrano, invece, sostanzialmente refrattarie a processi significativi di industrializzazione: la "quarta ondata" per ora non si vede» (da G. Becattini e G. Bianchi in «Note economiche», nn. 5-6, ed. Monte dei Paschi di Siena).

si di una entità completamente improduttiva e parassitaria, ma di una incarnazione infelice di quelli che Keynes chiamava gli «spiriti animali» del capitalismo, e dotata perciò di una sua perversa vitalità, la mafia imprenditrice ha rappresentato - e continua a rappresentare - una forma di utilizzo, sia pure limitata, delle forze della produzione. La stagnazione o il debole tasso di sviluppo economico delle zone mafiose risulta perciò dalla media aritmetica tra un alto saggio di crescita del settore illegale e delle imprese mafiose operanti sul mercato legale, e un saggio molto basso o negativo di crescita del settore legale e delle imprese non-mafiose.

La legge La Torre costituisce perciò una delle poche *chance* di autentico sviluppo economico della Sicilia e di buona parte del Mezzogiorno continentale. La libertà di mercato e la sicurezza di lungo periodo della produzione e degli investimenti garantita alle imprese non-mafiose dalla applicazione dura e senza tentennamenti della legge La Torre varrebbero molto di più - sullo stesso piano economico - di un ennesimo trasferimento clientelare di risorse dalla autorità centrale alla periferia. Vale forse la pena di concludere con la parafrasi di una nota affermazione di un vecchio economista liberale, Maffeo Pantaleoni: «Lo stato democratico è il quarto fattore della produzione».

Formazione di stelle e struttura delle galassie

Le osservazioni radio indicano che le nubi molecolari giganti nelle quali nascono le stelle sono distribuite nelle galassie a spirale in vario modo e questo forse spiega la diversità dell'aspetto ottico delle nubi

di Nick Scoville e Judith S. Young

Quasi tutte le stelle visibili nel cielo notturno si sono formate più di un miliardo di anni fa. La formazione di stelle è, però, un processo continuo e buona parte di quanto c'è di affascinante nell'universo è legato in un modo o nell'altro proprio al numero relativamente piccolo di stelle molto più giovani. Particolarmente interessanti sono i membri di massa maggiore e di vita più breve di ogni generazione di stelle: in gioventù cedono energia alle nebulose fluorescenti che costellano i bracci a spirale delle galassie come la nostra; al momento della morte esplodono come spettacolari supernove, colmando l'ambiente interstellare di una miscela di gas comprendente anche una frazione arricchita di elementi pesanti. È da queste ceneri che nasceranno le future generazioni di stelle. Negli ultimi anni si è capito, soprattutto grazie alle osservazioni radiotelescopiche, che la molla di questo processo di ringiovanimento è costituita da nubi molecolari giganti, estese per più di 100 anni luce e contenenti quantità di materiale gassoso pari anche a un milione di masse solari. In questi immensi bozzoli, in un'oscurità fredda e polverosa, compiono la loro metamorfosi le stelle.

Nonostante si sappia che le nubi giganti di gas sono regioni di intensa formazione stellare, è significativo che queste non si siano ancora completamente trasformate in stelle. Oggi infatti, a più di dieci miliardi di anni dalla nascita della Galassia, è possibile ancora vedere molte nuove stelle emergere dalle nubi in cui si sono formate. Isolate dall'ambiente galattico, le nubi collasserebbero sotto il loro stesso peso, trasformando in stelle in meno di dieci milioni di anni il proprio gas diffuso. Se la formazione di stelle fosse un processo inevitabile, dipendente solo dall'accumulo in una nube di una massa adeguata di materiale, la riserva di gas interstellare necessaria per la generazione successiva di stelle si sarebbe ridotta da tempo quasi a zero. Deve quindi esistere una qualche raffinata forma di interazione tra nubi e ambiente galattico, in grado di regolare efficacemente la formazione di stelle.

La percentuale di stelle giovani può variare notevolmente sia da una galassia all'altra sia da un luogo all'altro della stessa galassia. Nelle galassie a spirale come la nostra i bracci risultano piuttosto brillanti nelle fotografie eseguite in luce visibile perché in essi sono concentrate le stelle più giovani e massicce. Sebbene queste stelle abbiano una vita relativamente breve (meno di 10 milioni di anni), il loro irraggiamento può essere un milione di volte più intenso di quello del Sole. I luoghi dove sono nate risulteranno quindi fortemente illuminati per alcuni milioni di anni. Una delle caratteristiche più notevoli delle galassie a spirale, scoperta per la prima volta da Walter Baade del Mount Wilson Observatory quarant'anni fa, è la evidente correlazione tra le posizioni delle stelle massicce, la quale ci permette di percepire la grandiosa struttura a spirale. Spesso i bracci luminosi si possono seguire per un giro completo, corrispondente a una distanza che può anche raggiungere i 200 000 anni luce.

Come è possibile correlare la formazione di stelle massicce su tutto il disco galattico, una estensione che va molto oltre il raggio degli effetti fisici di una nube sull'altra o la sfera di influenza di una singola stella? Attualmente su questo problema esistono due scuole di pensiero. Chia-Chiao Lin e Frank H. Shu del Massachusetts Institute of Technology sono stati i primi ad avanzare l'ipotesi che queste strutture a grande scala siano onde di densità generate dalle interazioni gravitazionali collettive nel sistema galattico di miliardi di stelle o dalla forza di attrazione di una galassia vicina. Successivamente Philip E. Seiden e Humberto C. Gerola del Thomas J. Watson Research Center della IBM hanno proposto, in alternativa, che la formazione di stelle possa estendersi a tutta una galassia come l'incendio di un bosco: la formazione di stelle massicce in un punto scatenerebbe cioè quella di altre stelle nelle nubi vicine. (Una terza possibilità è che le nubi esistano e formino stelle in tutta la galassia, ma che in prossimità dei bracci a spirale abbiano proprietà diverse così che le stelle più massicce nascano soprattutto in queste zone.)

Da anni gli astronomi sono in grado di individuare in certe nubi di gas vicine al sistema solare altrettanti punti di attiva formazione di stelle. Fino a poco tempo fa, però, non era possibile osservare il fenomeno della nascita vera e propria. Una piccola quantità di polvere nelle nubi, circa l'1 per cento della loro massa, assorbe in maniera molto efficace la radiazione visibile e ultravioletta delle stelle giovani che ne sono avvolte tranne che nelle nubi più rarefatte. Queste microscopiche particelle di polvere sono composte di carbonio (in forma di grafite), silicati e altri composti simili ad alcune rocce terrestri e lunari. Nonostante che nello spazio interstellare la polvere sia molto dispersa, è facilmente visibile se si guarda la Via Lattea in una notte buia. In queste condizioni il disco galattico appare non come un'unica banda uniforme di stelle attraverso il cielo, ma come due bande separate da una striscia scura. Si vede questa spaccatura perché la luce delle stelle più lontane della Galassia viene assorbita dalle nubi di gas e di polvere più vicine a noi lungo la linea di vista.

Lo strato di polvere galattica che si vede proiettato contro il fondo di stelle ha uno spessore di 300 anni luce. A causa della polvere è impossibile osservare nel visibile le stelle estremamente giovani a meno che queste non si muovano uscendo dalla nube che le avvolge o non emettano abbastanza energia da disperderla. Invecchiando, le stelle interagiscono gravitazionalmente con le nubi interstellari massicce, aumentando gradualmente la distanza alla quale vengono scagliate fuori dal disco galattico prima di ricadervi. Di conseguenza, anche se probabilmente quasi tutte le stelle si sono formate da nubi di gas presenti nel disco sottile, lo spessore dello strato stellare più vecchio è salito a circa 1000 anni luce.

Una delle più note fucine di stelle della nostra galassia è la Grande Nebulosa di Orione, luminosa e facilmente visibile con un binocolo in mezzo alla spada sotto la Cintura di Orione. Oggi gli astronomi sanno che la sua emissione luminosa, studiata per la prima volta più di un secolo fa, è una manifestazione della fase finale del processo di formazione delle stelle. Vicino al centro di questa emissione si trova l'ammasso stellare del Trapezio, che contiene diverse stelle giovani di grande massa. La più massiccia irradia energia a un tasso 100 000 volte maggiore di quello del Sole. Una frazione significativa di questa radiazione si trova nella banda ultravioletta dello spettro, e possiede quindi abbastanza energia da strappare elettroni dall'idrogeno atomico, ionizzando così l'idrogeno contenuto nel gas circostante. L'intensa emissione visibile, dal caratteristico colore rosa, è in realtà luce fluorescente con una lunghezza d'onda di 656,3 nanometri emessa dall'idrogeno ionizzato quando si ricombina catturando un elettrone libero.

In tutta la nebulosa questo ciclo di ionizzazione e ricombinazione si ripete circa 10^{50} volte al secondo; per questo motivo la regione appare molto brillante nonostante si trovi a circa 1700 anni luce dal sistema solare. Nel gas della nebulosa la ionizzazione degli atomi di idrogeno procede a una velocità circa 1000 volte maggiore di quella del fenomeno inverso, ossia la ricombinazione di elettroni e di protoni che porta alla formazione di atomi di idrogeno. Si ha quindi una situazione di equilibrio quando la concentrazione di protoni e di elettroni è circa 1000 volte maggiore di quella dell'idrogeno neutro (non ionizzato). Le regioni come la nebulosa di Orione in cui gran parte del gas è ionizzato si chiamano regioni H II; si tratta delle nubi luminose che delineano i bracci a spirale delle galassie.

Essendo una delle regioni H II più vicine, la nebulosa di Orione è diventata la stele di Rosetta degli studi sulla formazione delle stelle e sui processi fisici connessi. Esaminando l'intera estensione della nebulosa visibile si notano sia agglome-

In questa fotografia ottenuta con il telescopio da quattro metri del Kitt Peak National Observatory si vede quasi frontalmente la galassia a spirale indicata con la sigla M101 nel catalogo Messier. La galassia, la cui distanza viene stimata in circa 20 milioni di anni luce, ha una estensione lungo la linea di vista di circa 300 000 anni luce. I bracci di spirale luminosi sono formati da regioni di idrogeno ionizzato, le regioni H II, rese fluorescenti dal calore delle stelle giovani di grande massa che contengono. I filamenti oscuri nei bracci sono strisce di polvere. Il nucleo relativamente piccolo e i bracci piuttosto aperti fanno di M101 un tipico esempio di galassia *Sc* *(si veda l'illustrazione a pagina 42).*

rati di materia stazionari, sia getti di plasma (gas ionizzato) ad alta velocità. Allontanandosi dall'ammasso del Trapezio si arriva alla fine al punto in cui tutti i fotoni ultravioletti emessi dalle quattro stelle del Trapezio si sono esauriti ionizzando il gas. Al di là di quel punto l'idrogeno non può più rimanere ionizzato. Fino a poco tempo fa la regione ionizzata era l'unica parte osservabile della nebulosa. Il forte oscuramento dovuto ai granelli di polvere mescolati al gas neutro impediva di vedere dentro la nube.

Oggi gli sviluppi dell'astronomia all'infrarosso e della radioastronomia a lunghezze d'onda millimetriche indicano che le osservazioni fatte precedentemente mostravano poco più della punta dell'iceberg. La radiazione infrarossa e quella millimetrica sono molto meno attenuate dalla polvere. La prima permette di individuare stelle profondamente immerse nella polvere, mentre la seconda dà modo di osservare il gas molecolare. I dati infrarossi sono in un certo senso il negativo dell'immagine ottica corrispondente, perché la polvere che assorbe la luce visibile riemetterà successivamente l'energia in eccesso alle lunghezze d'onda infrarosse. Nella regione infrarossa dello spettro, quindi, si osserva l'intensa emissione legata alle stelle molto luminose immerse nella nube di polvere; la stella in sé si vede di rado: di solito si tratta dell'energia riemessa dalla polvere circostante riscaldata. Le osservazioni della nebulosa di Orione nell'infrarosso hanno rivelato la presenza di due ammassi di stelle giovani profondamente immersi nella nube di idrogeno neutro. Il fatto che nessuno dei due fosse neppure debolmente visibile nelle immagini fotografiche della regione è insolito, tanto più che uno degli ammassi emette energia, sotto forma di radiazione infrarossa, a un tasso quasi pari a quello dell'ammasso del Trapezio, cioè circa 100 000 volte quello del Sole.

Entro la nube, dietro la nebulosa di Orione, il gas è abbastanza denso e freddo da consentire alla maggior parte degli atomi di combinarsi in molecole. In questa regione, dove la temperatura media è inferiore a 100 kelvin (gradi centigradi sopra lo zero assoluto), il componente più abbondante è l'idrogeno molecolare (H_2). Vi sono però anche numerosi composti in tracce, come il monossido di carbonio (CO), il cianogeno (CN) e l'ammoniaca (NH_3). Alle temperature normalmente esistenti nelle nubi l'idrogeno molecolare non è osservabile direttamente; per studiare questo gas freddo si è fatto uso delle molecole in tracce che, al contrario dell'H_2, sono in grado di assorbire e riemettere radiazioni alle lunghezze d'onda radio corte e nell'infrarosso lontano. Quando, 15 anni fa, Robert W. Wilson, Keith B. Jefferts e Arno A. Penzias dei Bell Laboratories rilevarono per la prima volta l'emissione del monossido di carbonio nella nebulosa di Orione, si aprì una fase particolarmente importante nello studio delle nubi dove si formano le stelle. Da allora, in queste regioni, sono state identificate circa 60 molecole, di complessità variabile fino ad arrivare al cianopentacetilene ($HC_{11}N$). L'elenco include l'acido formico (HCOOH), la formaldeide (H_2CO) e l'etanolo (C_2H_6O). Le molecole più complesse sono meno abbondanti del CO e possono essere osservate solo nel nucleo compatto delle nubi, dove la densità del gas è massima.

Il monossido di carbonio rimane comunque il migliore indicatore della presenza di gas molecolare su vaste regioni. Nella Via Lattea la densità media del gas interstellare è di circa un atomo per centimetro cubo. Le nubi molecolari, invece, che sono le zone relativamente dense del mezzo interstellare, hanno densità che variano di solito da molte centinaia a molte migliaia di molecole per centimetro cubo (un valore che è pur sempre solo un miliardesimo di miliardesimo della densità dell'atmosfera terrestre a livello del mare). Nelle nubi il monossido di carbonio è presente in ragione di circa una molecola ogni 10 000 molecole di idrogeno.

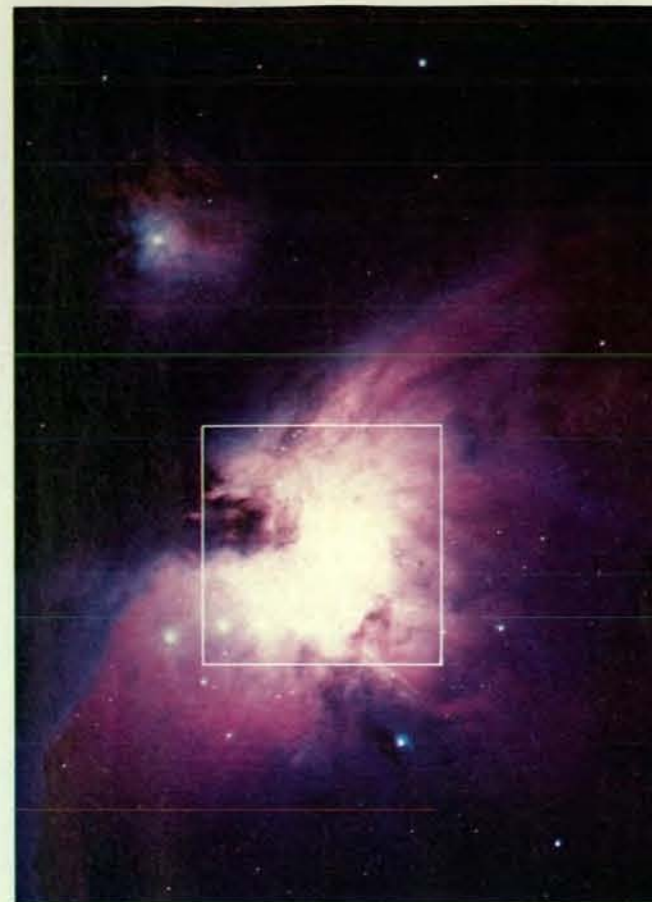
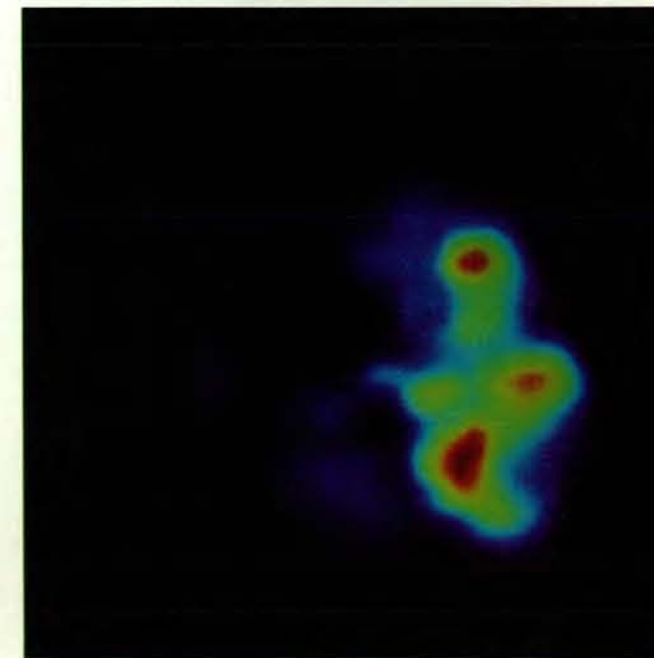
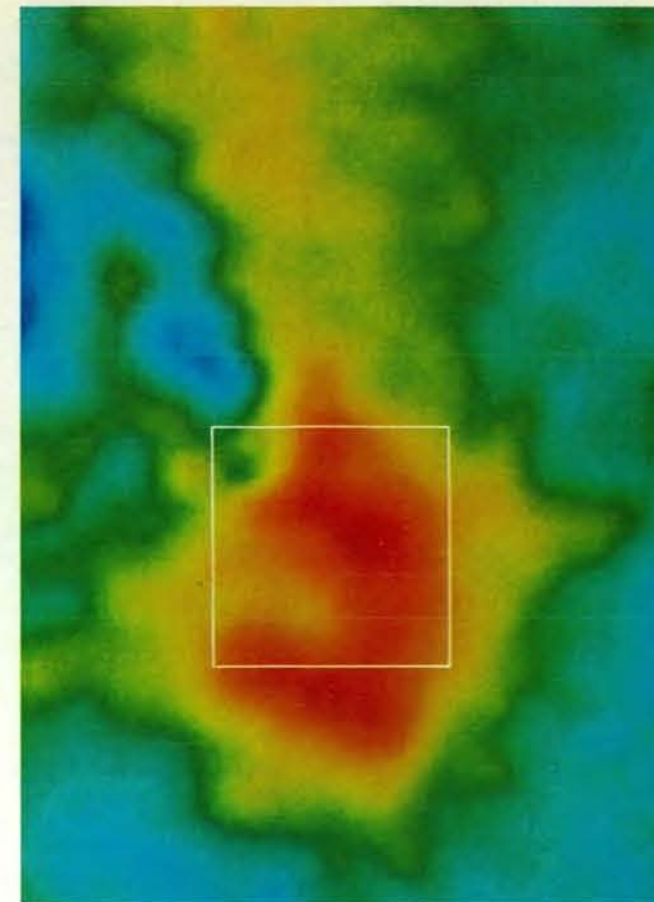
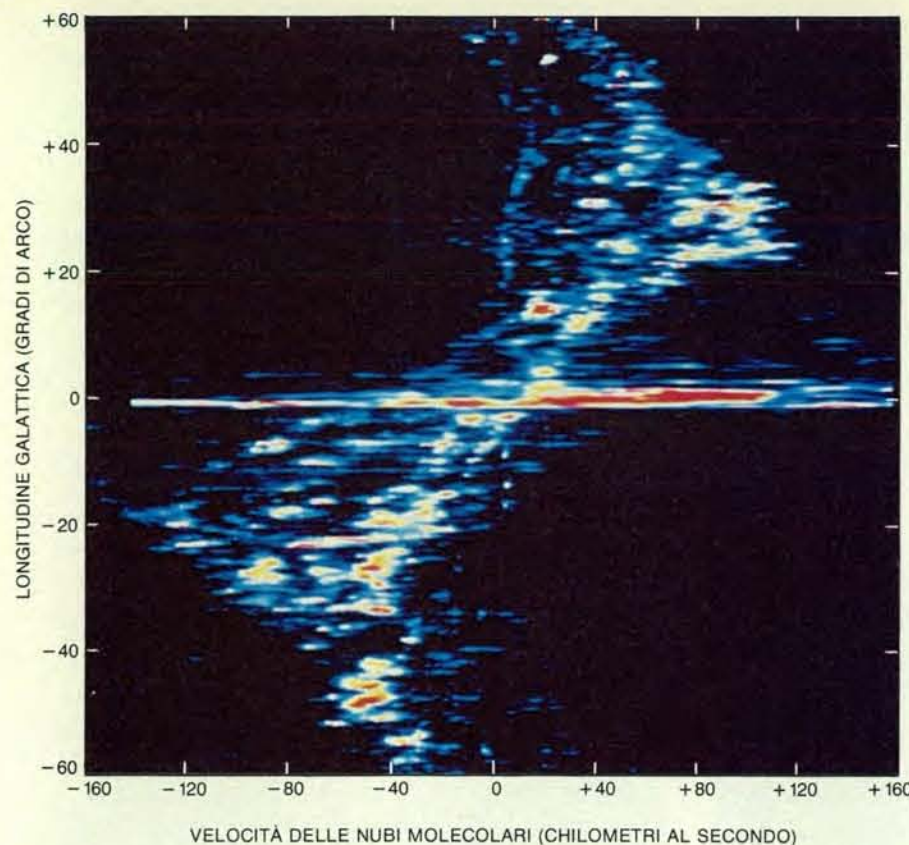
Una molecola come quella di ossido di carbonio si forma quando l'atomo dell'ossigeno e quello del carbonio condividono gli elettroni esterni, ciascuno dei quali trascorre così un certo tempo nelle vicinanze dell'altro atomo. La ripartizione degli elettroni non è esattamente equa e quindi a un'estremità della molecola si ha una piccola carica positiva netta e, all'altra estremità, una analoga carica negativa. È l'attrazione tra le cariche opposte che tiene gli atomi legati insieme.

La radiazione che proviene dalle molecole delle nubi interstellari deriva da mutamenti nella rotazione della molecola nel suo complesso; le rotazioni permesse non sono continue, ma quantizzate. Quando la molecola di monossido di carbonio passa da una velocità di rotazione elevata a una più bassa emette un fotone, un quanto di radiazione elettromagnetica, di energia pari alla diminuzione della sua energia rotazionale. Una transizione dal primo stato eccitato del monossido di carbonio a quello fondamentale produce un fotone alla lunghezza d'onda di 2,6 millimetri, pari a una frequenza radio di 115 000 megahertz. Poiché ogni specie molecolare ha una struttura leggermente diversa, ognuna irradia a una serie caratteristica di frequenze: le loro «impronte» spettrali sono distinte e identificabili, nella banda relativamente corta delle onde millimetriche.

L'apertura di questa regione dello spettro alle osservazioni astronomiche ha richiesto forti investimenti di risorse tecniche al National Radio Astronomy Observatory degli Stati Uniti, ai Bell Laboratories e alle numerose università in cui sono stati realizzati radioricevitori sensibili e telescopi con superfici riflettenti adeguatamente accurate. Sebbene all'inizio la tecnologia per le osservazioni radio alle lunghezze d'onda corte fosse onerosa, vi era un vantaggio collaterale: anche un telescopio relativamente piccolo fornisce a queste lunghezze d'onda un'ottima risoluzione angolare. Un telescopio da 14 metri, infatti, quando osserva la riga a 2,6 millimetri del monossido di carbonio ha una risoluzione di 50 secondi d'arco, mentre per ottenere una risoluzione analoga osservando la classica riga a 21 centimetri dell'idrogeno atomico occorrerebbe un telescopio con un diametro di circa un chilometro e mezzo.

Dall'emissione del monossido di carbonio osservata in una nube molecolare si possono ricavare non solo la densità e la temperatura delle molecole, ma anche i loro movimenti. Per rilevare i movimenti del gas lungo la linea di vista si misura lo spostamento Doppler della frequenza di emissione di un particolare volume di gas dalla frequenza della transizione misurata in laboratorio sulla Terra. La densità e la temperatura, invece, si ricavano in maniera meno diretta. In assenza di effetti esterni una molecola tende a emettere radiazioni e a decadere nello stato di minima energia rotazionale, in cui rimarrà finché l'ambiente non le fornirà abbastanza energia da rieccitarla. L'agente di eccitazione più importante è rappresentato dalle collisioni con l'idrogeno molecolare. Poiché la frequenza di collisione dipende dalla densità delle molecole, l'intensità della emissione molecolare di una nube dà una misura della densità dell'idrogeno molecolare. Nel caso del monossido di carbonio il decadimento radiativo è abbastanza lento rispetto ai tempi di collisione tipici perché la distribuzione di CO tra i vari stati rotazionali assuma una distribuzione termica caratteristica in tutte le nubi tranne quelle di densità minima. Nelle nubi più dense, quindi, l'intensità dell'emissione di CO fornisce una stima della temperatura dell'idrogeno molecolare. In queste regioni per misurare la densità del gas bisogna basarsi su altre molecole, come l'HCN (acido cianidrico) o il CS (monosolfuro di carbonio), che hanno un decadimento radiativo più veloce.

L'estensione della nube molecolare di Orione può essere valutata, osservando l'emissione del monossido di carbonio, più di tre gradi a sud e due gradi a nord della regione H II luminosa nel visibile. Questa misurazione angolare corrisponde a una distanza lineare di quasi 100 anni luce lungo l'asse maggiore. Le dimensioni lineari della nube molecolare sono quindi circa 50 volte maggiori di quelle della regione luminosa nel visibile. Attualmente si calcola che la quantità di gas allo stato molecolare presente nella regione sia di 200 000 masse solari, circa 1000 volte maggiore della massa totale delle stelle visibili nell'ammasso del Trapezio.

Nelle vicinanze della sorgente infrarossa più intensa della nebulosa la situazione non è affatto quella che ci si aspetterebbe per il collasso graduale di una nube fredda nel processo di formazione di stelle. Gli effetti delle emittenti infrarosse interne sono chiaramente visibili in un raggio di cinque anni luce: la loro radiazione riscalda la polvere circostante che a sua volta riscalda l'idrogeno molecolare. Le osservazioni del monossido di carbonio rivelano un gradiente termico di 80 gradi,

In questa pagina compaiono quattro immagini della nebulosa di Orione. La fotografia in alto a sinistra è stata ottenuta con il telescopio da quattro metri del Kitt Peak. L'immagine al calcolatore a falsi colori in alto a destra, che copre la stessa regione di cielo, è stata elaborata da una serie di dati sull'emissione di radioonde da parte delle molecole di monossido di carbonio nelle vicinanze della nebulosa, raccolti con il radiotelescopio da 14 metri del Five College Radio Astronomy Observatory del Massachusetts. La fotografia in basso a sinistra, ottenuta con il telescopio da tre metri del Lick Observatory, è una vista un po' più dettagliata della regione centrale della nebulosa, indicata dal riquadro nelle due immagini in alto. Questa fotografia, realizzata con un filtro che lascia passare soprattutto la luce verde emessa da atomi di ossigeno singolarmente ionizzati, rivela il gruppo centrale di stelle giovani luminose chiamato ammasso del Trapezio, responsabile della fluorescenza della nebulosa. L'immagine della stessa regione in basso a destra è stata ottenuta con il telescopio per l'infrarosso della NASA sul Mauna Kea, ad Hawaii: qui le chiazze luminose rappresentano le stelle più giovani della nebulosa, ancora immerse nella nube di polvere centrale, e quindi invisibili nell'ottico. I dati infrarossi sono stati raccolti da Gareth Wynn-Williams ed Eric E. Becklin dell'Università di Hawaii, Reinhart Genzel dell'Università della California a Berkeley e Dennis Downes dell'Istituto per la radioastronomia a onde millimetriche di Grenoble.

Mediante le onde radio a 2,6 millimetri emesse dalle molecole di monossido di carbonio delle regioni interne del disco galattico è stato effettuato un rilevamento delle nubi molecolari della nostra galassia. Si ritiene che il diametro delle singole nubi, rappresentate dalle strisce colorate, sia compreso tra 50 e 200 anni luce. Il primo rilevamento completo dell'interno della Galassia è stato effettuato riunendo i dati di una serie di misurazioni compiute sul piano galattico dal radiotelescopio di 14 metri del Massachusetts e da un radiotelescopio di quattro metri gestito dalla Commonwealth Scientific and Industrial Research Organization (CSIRO) in Australia. La scala verticale dà la longitudine galattica in gradi; lo zero corrisponde alla direzione del centro della Galassia. La scala orizzontale riporta la velocità delle singole nubi molecolari, ricavata dallo spostamento Doppler nella frequenza delle onde radio emesse rispetto alla frequenza normale di emissione del monossido di carbonio misurata in laboratorio. Le velocità osservate si possono attribuire principalmente al moto orbitale delle nubi attorno al centro galattico, quindi le misurazioni danno un'idea della posizione delle nubi nella Galassia. Le nubi sono concentrate soprattutto in due regioni: a longitudini galattiche inferiori a quattro gradi (che corrispondono a una regione di 2000 anni luce intorno al centro della Galassia) e a longitudini, sia positive sia negative, comprese tra 20 e 40 gradi (che corrispondono a un anello di materiale a metà della distanza tra il Sole e il centro galattico). Il rilevamento è stato compiuto da Daniel Clemens, David B. Sanders e uno degli autori (Scoville) dell'Università del Massachusetts ad Amherst, Philip M. Solomon della State University of New York a Stony Brook, Richard N. Manchester, Brian Robinson e John Whiteoak della CSIRO e William H. McCutcheon dell'Università della British Columbia. L'elaborazione dell'immagine a colori è dovuta al Remote Sensing Center dell'Università del Massachusetts.

dai 20 kelvin a cinque anni luce di distanza ai 100 kelvin a circa a 0,1 anni luce. A 0,1 anni luce avviene un brusco mutamento: il gas si muove a velocità supersoniche (fino a Mach 100) e una piccola percentuale dell'idrogeno molecolare viene riscaldata fino a oltre 2000 kelvin. Sebbene la causa di questo fenomeno non sia nota con precisione, si crede che i moti siano dovuti a una giovane stella vigorosa che si libera degli ultimi resti del bozzolo. Le alte temperature si troverebbero in corrispondenza dei fronti d'urto dove il gas supersonico espulso dalla giovane stella collide con la nube circostante. Può darsi che questi intensi flussi di materia, che rimescolano e sconvolgono di tanto in tanto la nube circostante, spieghino il tasso, generalmente basso, di formazione stellare nelle nubi. Forse la nascita di stelle sufficientemente potenti riesce davvero a limitare la successiva formazione di stelle nella stessa regione.

La nebulosa di Orione è solo una tra le tante regioni del genere della nostra galassia. Per determinare la quantità complessiva di gas molecolare nell'interno della Galassia uno di noi (Scoville), in collaborazione con Philip M. Solomon della State University of New York a Stony Brook, ha effettuato nel 1975 il primo campionamento dell'emissione del monossido di carbonio dal disco galattico, con il telescopio da 11 metri del National Radio Astronomy Observatory. I risultati erano molto diversi dalle precedenti immagini delle nubi di idrogeno atomico più rarefatte, chiamate regioni H I. Le nubi molecolari risultarono estremamente abbondanti in un raggio compreso tra 500 e 1000 anni luce di distanza dal nucleo galattico, mentre a distanze maggiori il loro numero calava sensibilmente. Più sorprendente fu la scoperta che la densità del gas molecolare saliva di nuovo raggiungendo un secondo picco, circa a metà della distanza tra il Sole e il centro galattico. Questo anello di gas molecolare con un picco a circa 15 000 anni luce dal centro della Galassia è stato osservato anche in successive e più complesse osservazioni dell'emissione del monossido di carbonio.

Poiché tutti i primi dati relativi alla Galassia sono stati raccolti con radiotelescopi situati nell'emisfero boreale, non si sapeva fino a che punto il piano galattico meridionale avesse la stessa struttura. Nel corso del 1983 un gruppo di radioastronomi guidato da Brian Robinson della Commonwealth Scientific and Industrial Research Organization (CSIRO) australiana ha condotto a termine uno studio sull'emissione del monossido di carbonio nell'emisfero australe. I quantitativi di gas molecolare osservati in zone equivalenti del lato nord e del lato sud della Galassia sono in accordo entro il 20 per cento. La distribuzione del gas a sud presenta, come quella a nord, un picco anulare a metà strada tra il Sole e il centro galattico, ma la forma della parte meridionale dell'anello è un po' diversa: il picco ha una densità inferiore del 30 per cento e una ampiezza del 30 per cento maggiore.

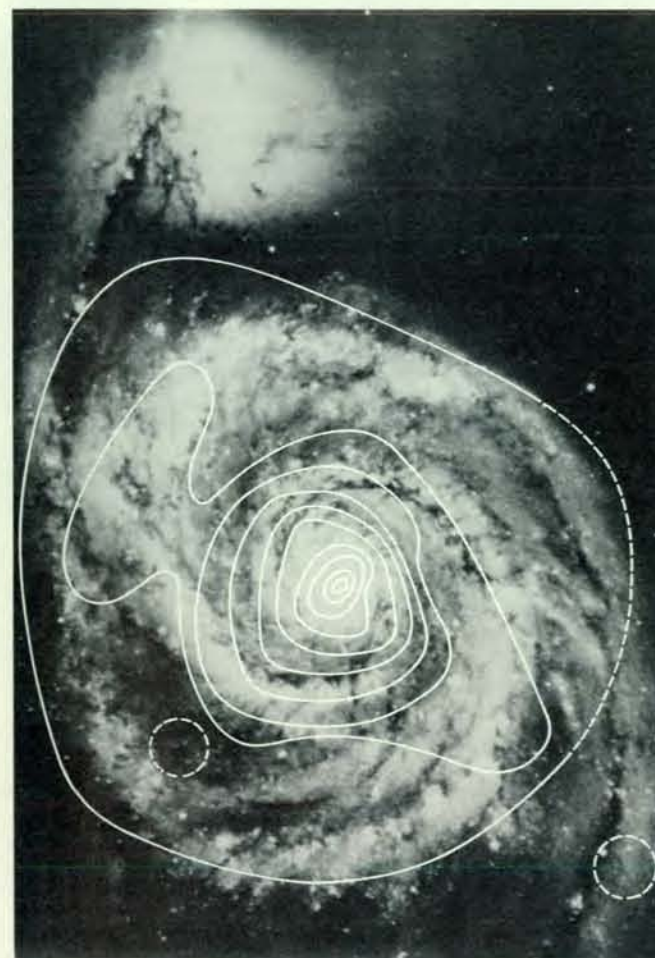
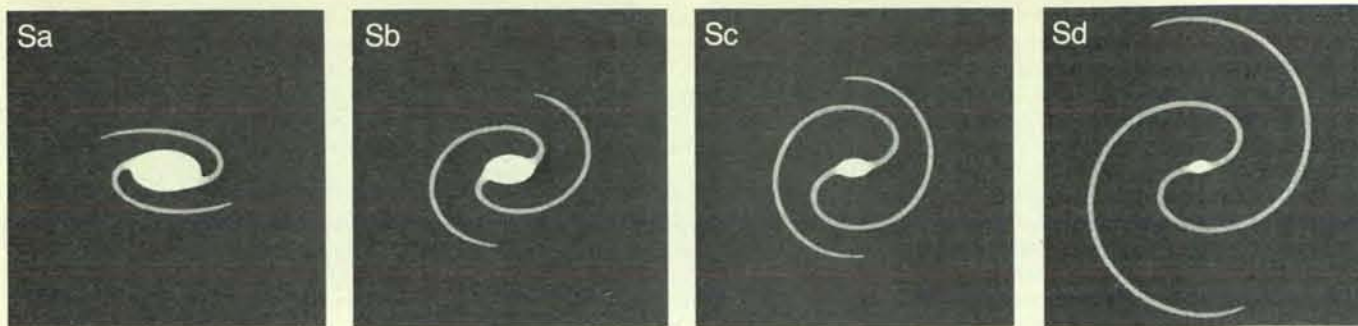
La massa totale di gas molecolare stimata sulla base dell'emissione del monossido di carbonio all'interno della Galassia è compresa tra uno e tre miliardi di masse solari, equivalenti a circa il 15 per cento della massa stellare totale nella stessa regione. Questa quantità di gas è di gran lunga superiore a quella del gas atomico interstellare presente nella regione ed è in realtà paragonabile alla quantità di idrogeno atomico presente nell'intera Galassia fino a una distanza doppia di quella del Sole dal centro galattico. Va ricordato che fino a dieci anni fa questo importante componente della nostra galassia era completamente ignoto su scala galattica. L'importanza di questo gas deriva non solo dalla sua abbondanza rispetto ad altre forme di materia, ma anche dal fatto che sono le nubi di idrogeno molecolare, e non quelle di idrogeno atomico, il mezzo da cui si formano le nuove stelle.

Le proprietà delle nubi molecolari sono le più importanti. La nebulosa di Orione infatti, lungi dall'essere eccezionalmente grande, è una delle più piccole tra le nubi molecolari giganti della Galassia. Nel 1981 David B. Sanders, che allora studiava a Stony Brook, misurò le dimensioni di più di 300 nubi dell'anello galattico, trovando che la maggior parte del gas era contenuta in nubi con un diametro medio di circa 100 anni luce. Nonostante che le nubi siano estremamente rarefatte (circa 300 molecole di idrogeno per centimetro cubo), il loro volume è così grande che la loro massa totale può variare da 100 000 a diversi milioni di volte la massa del Sole. In realtà oggi si pensa che le nubi molecolari giganti siano gli oggetti più massicci di tutta la Galassia. E le nubi con un diametro superiore ai 50 anni luce sono probabilmente circa 5000.

I rapporti tra le nubi molecolari giganti e gli altri componenti della Galassia (le stelle giovani massicce, il gas più diffuso e le stelle più vecchie), possono fornire un importante contributo per la comprensione della vasta macchina galattica. Le nubi molecolari giganti si formano per collisione di molte nubi molecolari più piccole o per compressione di idrogeno atomico diffuso? C'è sempre una nube molecolare vicino a una nuova stella e, se è così, i luoghi di nascita delle stelle sono vicini alla superficie della nube o in profondità? La prima possibilità potrebbe aver luogo se il collasso delle nubi fosse attivato da fattori esterni, come la collisione tra una nube e l'altra.

Le posizioni di stelle giovani di grande massa come quelle del Trapezio si possono individuare con precisione in tutta la Galassia grazie all'emissione radio a bassa frequenza prodotta dal gas ionizzato che le circonda. Come ci si aspetterebbe c'è una buona correlazione tra le regioni H II e le nubi: praticamente tutte le regioni H II che si conoscono nel radio o nell'ottico, di dimensioni equivalenti a quelle della nebulosa di Orione, hanno nelle loro vicinanze una nube molecolare. Di solito a queste regioni H II corrispondono le nubi più grandi e più calde. Tenendo conto del fatto che sono proprio le regioni H II a delineare meglio i bracci a spirale delle galassie esterne, questa correlazione fa pensare che la maggior parte delle nubi giganti si trovi anche lungo i bracci della nostra galassia. In realtà nella nostra galassia le nubi molecolari giganti sono più numerose delle regioni H II giganti conosciute (5000 contro 200 circa), ed evidentemente molte delle nubi più piccole e più fredde non si trovano vicino ad alcuna regione H II. Per questo bisogna essere cauti nell'interpretare le osservazioni relative al monossido di carbonio eseguite a bassa sensibilità o con una scarsa risoluzione angolare. I dati ottenuti in questo modo fanno facilmente pensare a bracci a spirale, ma solo perché le osservazioni sono inerentemente viziate e favorirebbero le nubi più grandi e più calde anche qualora la distribuzione delle nubi, per dimensioni e temperatura, fosse abbastanza uniforme.

Secondo un modello proposto recentemente da John Y. Kwan dei Bell Laboratories e dell'Università del Massachusetts ad Amherst sarebbe possibile spiegare l'esistenza di nubi particolarmente grandi nei bracci se quelle più piccole delle zone tra i bracci collidessero e si aggregassero in poche supernubi, cioè ammassi di nubi, quando raggiungono i bracci. Se la frequenza delle collisioni tra le nubi fosse più elevata nei bracci, si potrebbe spiegare il maggiore tasso di formazione di stelle massicce in queste zone. Quando, infatti, due nubi collidono, la compressione del gas all'interfaccia dovrebbe essere seguita dal collasso di un frammento di nube in una o più protostelle. Queste regioni compresse risulterebbero adatte alla formazione di stelle perché la autogravità del frammento di nube aumenterebbe a causa della maggiore densità.

Una volta che le stelle massicce si sono formate, la loro elevata luminosità riscal-

La distribuzione delle nubi molecolari in M51, la galassia Vortice *(a sinistra)*, è strettamente legata a quella della radiazione nell'infrarosso lontano dovuta a stelle giovani *(a destra)*. In tutti e due i casi i profili indicano un aumento di concentrazione verso il centro galattico. La somiglianza tra le due distribuzioni fa pensare che le stelle si formino nelle nubi e che il loro tasso di formazione sia direttamente proporzionale alla quantità di gas molecolare in ogni punto. La mappa delle nubi molecolari si basa su dati relativi al monossido di carbonio ottenuti dagli autori con il radiotelescopio da 14 metri del Massachusetts. La mappa nell'infrarosso lontano è stata realizzata da James Smith e collaboratori con l'osservatorio aereo C-141 della NASA. Si noti la mancanza di una chiara correlazione con i bracci di spirale della galassia.

Le galassie a spirale vanno dal tipo *Sa*, con un grosso rigonfiamento centrale e una struttura a bracci molto stretti, al tipo *Sd*, che ha una struttura a bracci aperti e un nucleo centrale piuttosto piccolo. Si ritiene che la nostra galassia abbia una struttura intermedia, di tipo *Sbc*.

da la nube di polvere circostante; per questo la corrispondenza tra nubi calde e posizioni delle regioni H II è comprensibile. Dopo aver lasciato i bracci, le supernubi potrebbero suddividersi in unità più piccole, forse per le forze perturbatrici esercitate dalle regioni H II calde.

L'osservazione del gas molecolare così diffuso nella nostra galassia solleva quesiti fondamentali a cui è possibile rispondere solo osservando altre galassie. Per esempio, l'anello di nubi molecolari situato a 15 000 anni luce dal centro galattico è una caratteristica comune ad altre galassie? Si tratta di una traccia della forma originaria della Galassia o la Galassia ha assunto questa configurazione nel corso della vita? La relativa abbondanza di gas molecolare osservato nell'interno della nostra galassia è una caratteristica generale della maggior parte delle galassie a spirale? La quantità e la distribuzione delle molecole in una galassia dipendono dalla sua forma? E infine, come dipende la luminosità totale di una data galassia dall'entità e dalla distribuzione delle nubi molecolari? Vi dovrebbe essere una certa correlazione se gran parte dell'energia di una galassia fosse generata dalle giovani stelle che si formano entro le nubi.

Oggi gli studi nel visibile hanno stabilito con sicurezza che le proprietà delle stelle delle varie galassie seguono andamenti prevedibili. Nel suo lavoro pionieristico sulle galassie Baade osservò che è possibile dividere le stelle in due classi, la prima comprendente stelle giovani blu e la seconda stelle vecchie e rosse, che chiamò rispettivamente Popolazione I e Popolazione II.

Nelle galassie ellittiche e lenticolari le stelle hanno quasi tutte più di cinque miliardi di anni. Tranne che in pochi casi non si vede traccia né di stelle giovani, di Popolazione I, né di quantità significative di gas interstellare disponibile per la formazione di future generazioni di stelle; in queste galassie la riserva di gas o si è esaurita molto tempo fa o è stata spazzata via nello spazio intergalattico circostante. Queste galassie contengono soprattutto stelle di Popolazione II.

Nelle galassie a spirale come la nostra, invece, sono presenti entrambe le popolazioni. Le stelle vecchie della Popolazione II occupano un volume sferoidale simile a quello di una galassia ellittica, mentre quelle giovani della Popolazione I si trovano quasi solo nel disco sottile insieme a praticamente tutta la restante materia interstellare. La distribuzione sferoidale di stelle vecchie e il disco più giovane sono concentrici. Tra le galassie a spirale il rapporto tra la componente giovane e quella vecchia piò variare in misura notevole andando dal primo tipo di spirali indicate con *Sa*, che hanno un rigonfiamento centrale molto grande e bracci a spirale avvolti strettamente, all'ultimo tipo *Sd*, con un nucleo visibile quasi insignificante e una struttura a spirale molto aperta e irregolare. Si ritiene che la nostra galassia sia del tipo *Sbc*, con caratteristiche intermedie. All'interno di ogni classe morfologica si incontra una vasta gamma di masse e di luminosità, estesa di solito per un fattore compreso tra 10 e 100 in termini di energia totale emessa.

I primi studi sulla radiazione emessa dalle molecole in altre galassie sono stati compiuti da Lee J. Rickard della Howard University e da Patrick Palmer dell'Università di Chicago, che inizialmente concentrarono le ricerche su galassie con un'emissione infrarossa eccezionalmente intensa. Si credeva, infatti, che in queste galassie un'improvvisa proliferazione (*burst*) di stelle fosse alimentata da una ricchissima riserva di gas molecolare. Negli ultimi tempi abbiamo intrapreso un programma articolato per la realizzazione di mappe relative all'emissione di monossido di carbonio in galassie a spirale più normali, al fine di chiarire la relazione tra il contenuto e la distribuzione del gas molecolare da una parte e la morfologia e la luminosità delle galassie dall'altra. Abbiamo effettuato le osservazioni con il radiotelescopio da 14 metri gestito dal Five College Radio Astronomy Observatory del Massachusetts; è il più grande strumento del tipo negli Stati Uniti e fornisce una elevata risoluzione angolare, che ci ha permesso di osservare anche piccoli dettagli delle galassie esterne. Delle 80 galassie che abbiamo esaminato, quasi metà hanno rivelato un'emissione di monossido di carbonio misurabile e abbiamo realizzato mappe parziali di 20 di esse. La maggior parte di queste galassie sono classificate tra le spirali normali, anche se alcune sono irregolari.

Date le grandi distanze delle galassie esterne non è possibile osservare in esse singole nubi molecolari. La risoluzione del radiotelescopio da 14 metri (50 secondi di arco) ci permette però di esaminare l'emissione composita di regioni che si estendono di norma 5000 anni luce e che contengono diverse nubi molecolari. Le osservazioni relative al monossido di carbonio nelle galassie esterne ci consentono quindi di determinare la distribuzione globale di queste nubi, ma non le loro proprietà individuali.

Una delle galassie a spirale maggiormente studiata da diversi gruppi di radioastronomi è M51, la galassia Vortice. In questo caso l'emissione del monossido di carbonio è stata rilevata sull'intero disco visibile e, come in molte altre galassie, la concentrazione massima di gas molecolare si osserva vicino al centro. Un fatto molto sorprendente, però, è l'esistenza di una diminuzione di concentrazione abbastanza sistematica e uniforme dal centro verso l'esterno fino al limite dove l'emissione diventa impercettibile. In altri termini non c'è traccia di concentrazioni del gas molecolare né in un anello, né lungo i bracci. L'assenza di concentrazioni di questo genere è in parte attribuibile all'insufficienza della risoluzione, perché i bracci sono molto sottili e ravvicinati. L'assenza di un anello, invece, è pienamente significativa: se esistesse, una struttura simile a quella della nostra galassia sarebbe facilmente osservabile.

Da un confronto eseguito in M51 tra la distribuzione della materia allo stato molecolare e la distribuzione di luminosità, soprattutto della luminosità dovuta a stelle giovani, è possibile ricavare un elemento importante per capire come il tasso di formazione delle stelle vari da un punto all'altro del disco galattico. Con l'aiuto di un telescopio a bordo dell'osservatorio aereo C-141 della National Aeronautics and Space Administration, James Smith dello Yerkes Observatory ha eseguito recentemente una mappa completa di M51 nell'infrarosso lontano a lunghezze d'onda comprese tra 80 e 200 micrometri. Alla radiazione in questa regione dello spettro contribuiscono anche sorgenti simili alle intense emittenti infrarosse che si trovano dietro la nebulosa di Orione, le quali presumibilmente sono giovani ammassi di stelle formatisi negli ultimi dieci milioni di anni e ancora avvolti dalla polvere. La luminosità totale misurata da Smith nell'infrarosso lontano è pari a 30 miliardi di unità solari entro la regione del disco visibile fino a un raggio galattico analogo alla distanza che separa il Sole dal centro della nostra galassia.

È di fondamentale importanza la scoperta che esiste una proporzionalità quasi perfetta tra la luminosità nell'infrarosso lontano e l'intensità di emissione del monossido di carbonio in diversi punti di M51. Tutte e due infatti decrescono regolarmente con la distanza dal centro della galassia e la loro dipendenza dal raggio è quasi identica. Se il tasso di formazione delle stelle è rispecchiato fedelmente dall'emissione di energia nella regione infrarossa e se, analogamente, l'intensità dell'emissione del monossido di carbonio dà una misura della disponibilità di gas in grado di formare stelle, allora si potrebbe dedurre che il tasso di formazione delle stelle dipende solamente dall'abbondanza delle nubi molecolari e non dalla loro posizione nella galassia.

Sulle prime questa conclusione sembra sorprendente, perché le forze esterne che promuovono il collasso delle nubi e quindi la formazione delle stelle dovrebbero dipendere strettamente dalla distanza dal centro della galassia. Forse una spiegazione naturale della semplice correlazione tra il tasso di formazione delle stelle e la massa di materia in forma molecolare va cercata nella natura delle nubi. Se le nubi di M51 sono soprattutto nubi giganti come quelle della nostra galassia, è difficile che fenomeni esterni come i gusci in espansione delle supernove e le regioni H II riescano a penetrarvi molto in profondità. L'inerzia di una nube con una massa pari a un milione di masse solari è semplicemente troppo grande perché questi agenti ne influenzino una parte significativa. Queste nubi sono già «gravide» di stelle in formazione e uno stimolo esterno che superi solo gli strati superficiali non modifica il tasso complessivo di formazione di stelle dell'intera nube.

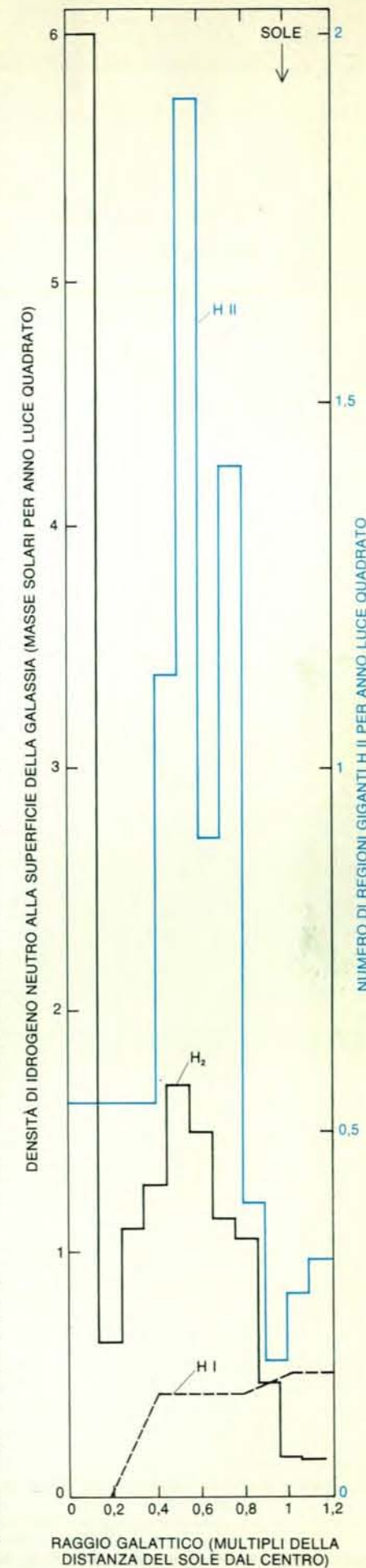
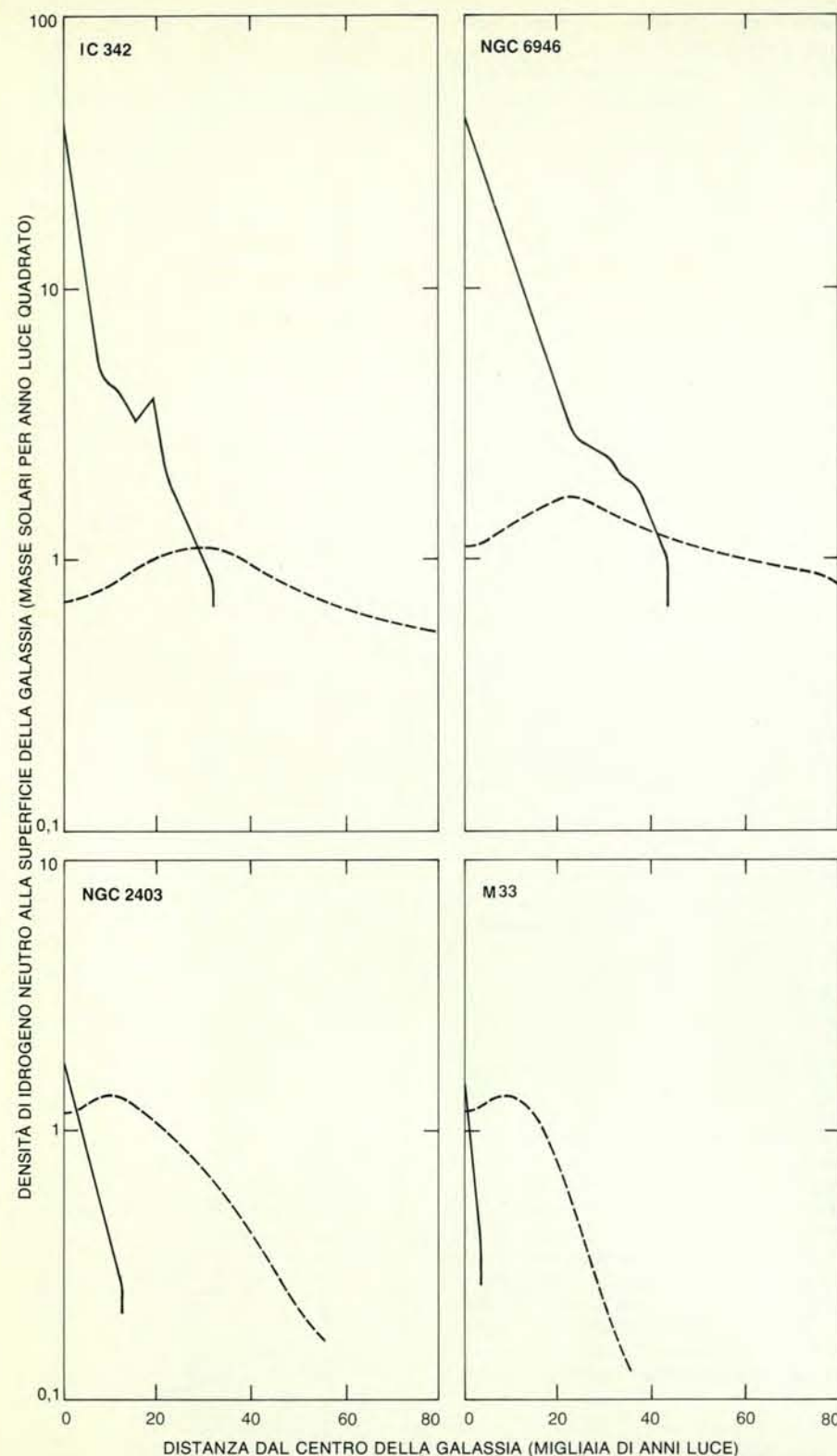
Oggi sembra che la proporzionalità tra il tasso di formazione delle stelle e l'abbondanza del gas molecolare individuata in M51 sia una regola generale in quelle galassie a spirale di elevata luminosità dell'ultimo tipo dove la quantità di gas molecolare supera quella del gas atomico. Per la maggior parte delle galassie non esistono ancora dati infrarossi completi, ma si può ottenere una valutazione approssimativa del tasso di formazione delle stelle in base alla componente blu della luce stellare del disco galattico. Questa luce è prodotta soprattutto da stelle calde di età inferiore a qualche miliardo di anni. Nelle spirali IC 342 e NGC 6946 (di tipo *Scd*) abbiamo trovato analoghe variazioni con il raggio dell'emissione del monossido di carbonio e della luce blu.

Dal momento che l'emissione del monossido di carbonio indica la distribuzione della componente più densa del mezzo interstellare, è interessante confrontarla con l'emissione dell'idrogeno atomico, che presumibilmente individua il gas meno denso (ma sempre neutro). Negli ultimi 10-20 anni molti astronomi hanno compiuto studi dettagliati del contenuto di idrogeno atomico delle galassie esterne e della sua distribuzione. Nonostante la loro vasta gamma di luminosità le galassie *Scd* hanno tutte distribuzioni simili di idrogeno atomico, e precisamente una densità abbastanza costante su tutto il disco tranne al centro, dove c'è una carenza di H I.

In generale per ogni galassia la dimensione dell'inviluppo di idrogeno atomico è molto più grande anche della galassia visibile. Nelle stesse galassie il contenuto di monossido di carbonio e la sua distribuzione sono invece molto diversi e assomigliano ben poco ai profili dell'idrogeno atomico. In particolare, si è scoperto che le galassie di luminosità elevata sono ricche di nubi molecolari, mentre quelle di bassa luminosità contengono solo piccoli quantitativi di gas molecolare. Inoltre le distribuzioni di questo gas mostrano un elevato gradiente in senso radiale. Quindi le galassie a elevata luminosità contengono soprattutto gas molecolare al centro e gas atomico nelle regioni esterne, mentre quelle meno luminose sono più ricche di gas atomico dappertutto. Sotto questo aspetto la nostra galassia assomiglia maggiormente alle galassie esterne a elevata luminosità.

L'ipotesi di fondo, naturalmente, è che osservando le altre galassie si arrivi a comprendere le caratteristiche della nostra. Dato che la nostra galassia occupa una posizione intermedia nella classificazione delle galassie a spirale, all'inizio è parso moltro strano che nessuna delle prime galassie studiate per il loro contenuto di monossido di carbonio mostrasse un picco centrale e un anello di nubi molecolari come quelli della nostra. Oggi, però, è chiaro che questi risultati erano almeno in parte da attribuire a un effetto di selezione osservativa: la maggior parte di queste prime galassie osservate erano infatti polverose spirali *Sc*, non di tipo intermedio come la nostra. Man mano che si osservano galassie *Sa* e *Sb*, invece, l'anello è sempre più frequente. La sua origine forse è legata alle dimensioni del rigonfiamento centrale di stelle vecchie.

In realtà, confrontando la distribuzione del gas molecolare in tutte le galassie a spirale il disco esterno presenta in tutti i

Questo diagramma riporta gli indizi dell'esistenza nella nostra galassia di un anello di materiale in cui si formano stelle. La curva in colore indica la distribuzione delle regioni giganti di idrogeno ionizzato (H II), un buon indicatore delle zone in cui si stanno formando stelle di grande massa. La curva in nero indica invece la distribuzione dell'idrogeno molecolare (H_2). Entrambe le distribuzioni presentano un netto picco a una distanza dal centro galattico pari a circa metà di quella del Sole. La concentrazione, invece, dell'idrogeno atomico neutro, ossia non ionizzato (H I), indicata dalla curva tratteggiata in nero, rimane abbastanza costante, aumentando un po' oltre l'orbita del Sole, dove vi sono alcune regioni H II giganti.

La composizione della materia interstellare in quattro galassie di tipo analogo risulta strettamente legata alla loro luminosità. Tutte queste galassie sono classificate di tipo *Scd*: spirali aperte con un nucleo relativamente piccolo. Vicino al centro delle due galassie più luminose, IC 342 *(in alto a sinistra)* e NGC 6946 *(in alto a destra)* il rapporto tra idrogeno molecolare H_2 *(curva in nero)* e idrogeno atomico neutro H I *(curva tratteggiata)* è di circa 100 a 1, mentre nelle galassie meno luminose, NGC 2403 *(in basso a sinistra)* e M33 *(in basso a destra)* le quantità di idrogeno molecolare e atomico sono circa uguali. Le quattro galassie hanno valori analoghi per quanto riguarda l'idrogeno neutro in superficie, cosicché è il quantitativo di gas molecolare che varia dall'una all'altra. In altri termini le galassie più luminose contengono molto più gas molecolare di quelle meno luminose. Le galassie meno luminose sono anche più piccole, come indica l'estensione limitata della regione in esse occupata dall'idrogeno atomico. Le osservazioni relative all'idrogeno molecolare sono state eseguite dagli autori; quelle relative all'idrogeno atomico sono state eseguite da David H. Rogstad e da G. Seth Shostak del California Institute of Technology.

casi strutture abbastanza simili, con una diminuzione uniforme di concentrazione; è nelle regioni interne del disco che si trovano le differenze. In alcune galassie vi è una brusca diminuzione della densità di idrogeno molecolare al centro, mentre in altre la concentrazione cresce regolarmente dal disco esterno fino al nucleo. Questa osservazione fa pensare che la caratteristica significativa non sia il picco dell'anello, ma il vuoto che a volte compare nelle regioni interne. Potrebbe forse esserci un legame tra le dimensioni della popolazione stellare del rigonfiamento centrale e l'esistenza di una lacuna nella distribuzione di gas se il gas originariamente presente in questa zona si fosse esaurito molto tempo fa per formare le stelle del massiccio rigonfiamento centrale. Perché alcune galassie formino un nucleo più massiccio rimane però un mistero legato alla loro costituzione iniziale.

La relazione generale tra la distribuzione del monossido di carbonio e la luminosità ottica, che avevamo scoperto in origine in poche galassie selezionate fra quelle dell'ultimo tipo, vale anche se si considera un campione di galassie su una vasta gamma di luminosità, ma tutte di una data classe morfologica. I risultati per le regioni centrali delle galassie *Sc* sono decisamente notevoli: la luminosità ottica è direttamente proporzionale alla quantità di gas molecolare. In altri termini, se vi sono più nubi molecolari si formeranno più stelle, e la galassia avrà di conseguenza una luminosità più elevata. Questa relazione è esattamente la stessa che si osserva nelle singole galassie *Sc* come M51, in cui si era trovato che la distribuzione dell'emissione del monossido di carbonio riproduce quella delle stelle giovani. L'esistenza della correlazione sia all'interno di determinate galassie *Sc*, sia tra galassie diverse, indica però che si tratta di una caratteristica generale della formazione di stelle nelle galassie di questo tipo. Infatti se nello stesso campione di galassie si confrontassero i contenuti di idrogeno atomico al centro con la luminosità ottica non si troverebbe nessuna correlazione.

Le galassie più enigmatiche sono forse quel 10 per cento circa le cui regioni centrali presentano un'intensa attività sotto forma di emissione X o radio, moti decisamente non circolari del gas e forte irraggiamento ottico e infrarosso. A un livello di attività molto basso anche il centro della nostra galassia rivela buona parte di questi sintomi. Nei casi più estremi, come i quasar, si crede che l'attività sia prodotta da un oggetto straordinario, per esempio un buco nero di grande massa al centro della galassia. In altri casi è stata avanzata l'ipotesi che l'attività si possa attribuire a una improvvisa proliferazione di stelle particolarmente intensa, una vera e propria «esplosione». In quest'ultimo caso le giovani stelle massicce (fino a un miliardo di masse solari nei casi più eccezionali) avranno una luminosità enorme per un periodo breve di circa 10 milioni di anni; per questo una improvvisa proliferazione di stelle può spiegare una elevata e istantanea emissione di energia. Per di più le esplosioni di supernova che si verificherebbero verso la fine dell'improvvisa proliferazione di stelle genererebbero moti di gas di velocità elevata oltre a una forte emissione X e radio. Chiaramente la durata di una proliferazione improvvisa è limitata dall'entità della riserva di materia interstellare vicino al centro della galassia, perché alla fine quasi tutto il gas condensa in stelle e il tasso di formazione diminuisce.

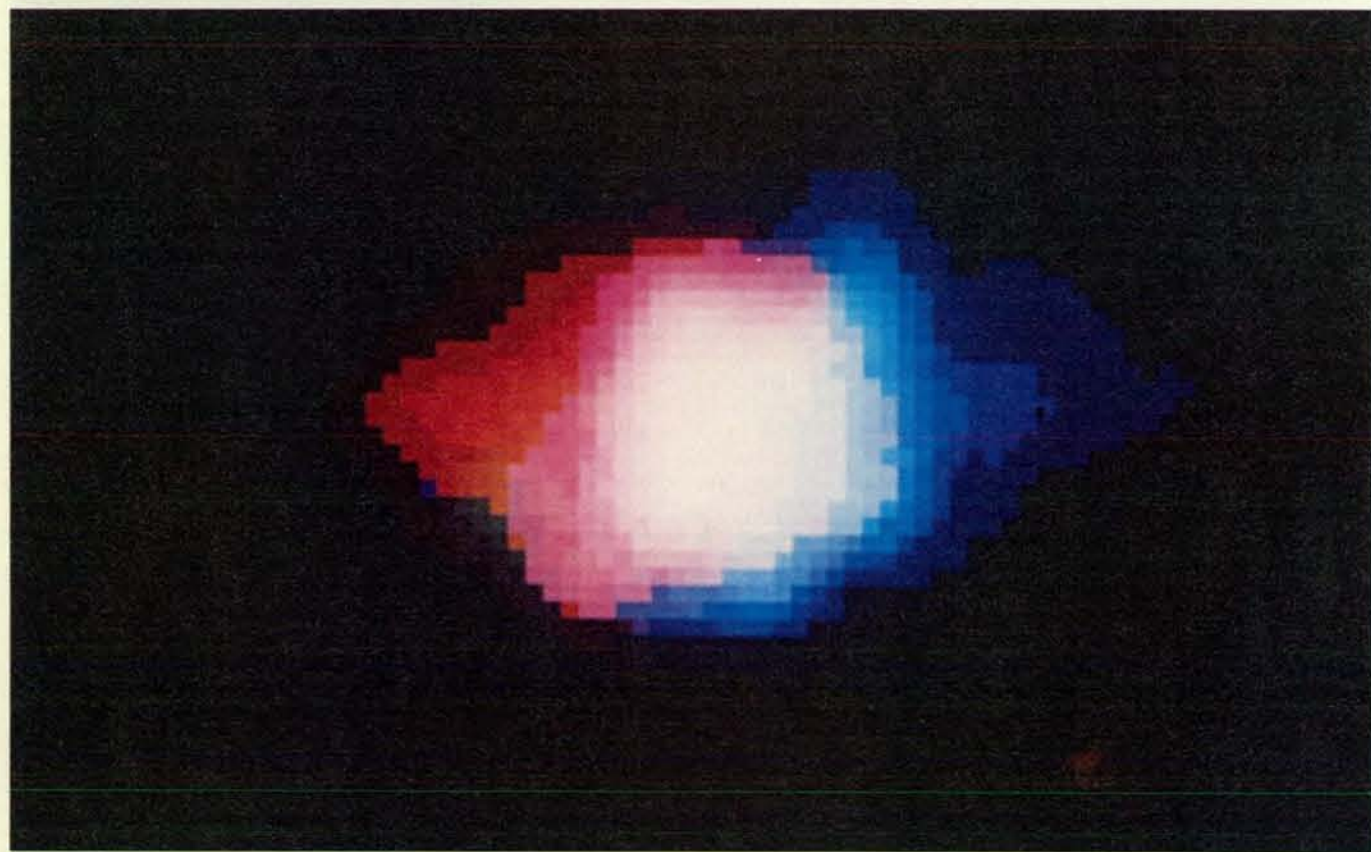
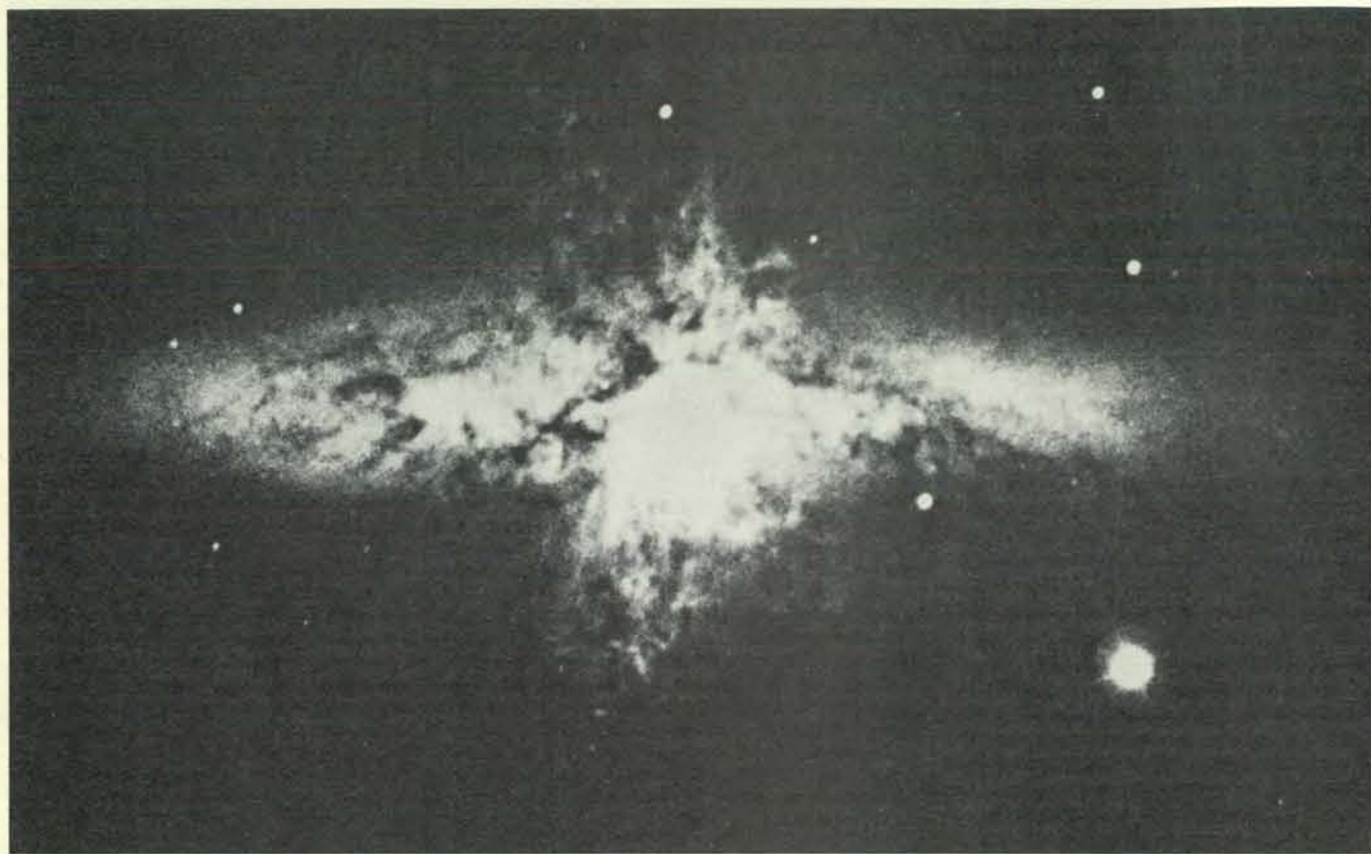
Uno degli esempi più vicini a noi di questo tipo di fenomeni è la galassia irregolare M82. Per molti anni si è creduto che il nucleo di questa galassia fosse esploso, perché sopra e sotto il disco si potevano osservare pennacchi di gas ionizzato ad alta velocità. I filamenti si dipartono dal centro in direzione radiale proprio come se si trattasse di un getto di gas diretto verso l'esterno. Oggi invece l'opinione più diffusa è che molte delle stranezze che si vedono intorno al nucleo di M82 vadano attribuite non a gas che si allontana dalla galassia, ma a gas intergalattico che vi cade dentro. Il ponte di idrogeno che si estende nel cielo fino alla vicina galassia a spirale M81 fa pensare che questo gas sia stato strappato all'alone esterno di M81 durante un passaggio ravvicinato.

M82 è una delle più luminose sorgenti infrarosse al di fuori della nostra galassia. La sua luminosità infrarossa al centro è circa venti miliardi di volte superiore a quella del Sole. Data l'abbondanza di polvere opaca nel piano della galassia la luminosità nel visibile della stessa regione è circa venti volte inferiore. Come era prevedibile M82 è risultata una delle più forti sorgenti di emissione del monossido di carbonio fuori dalla nostra galassia. È stata infatti la prima galassia studiata da Rickard e collaboratori relativamente alla emissione di monossido di carbonio. Un'analisi dei moti del gas molecolare condotta sulla base dell'effetto Doppler su questa radiazione rivela una disposizione con una struttura ben precisa, in cui il gas a nord si allontana da noi e quello a sud si avvicina. Contrariamente poi a quanto succede nelle galassie a spirale normali, il cui asse di rotazione è perpendicolare al disco galattico, in questo caso si trova che l'asse è inclinato di quasi 45 gradi rispetto al disco. Questa scoperta suggerisce l'esistenza di forti spostamenti in direzione radiale, verso il centro di M82 o in direzione opposta. In altre parole le nubi di gas hanno velocità decisamente diverse da quelle delle comuni orbite circolari. Questi moti così particolari sono dovuti forse a una caduta di gas dall'esterno della galassia.

Nel nucleo della galassia a spirale IC 342 le nubi molecolari sono raccolte in una regione a forma di barra lunga circa 6000 anni luce *(profili in nero)*. I dati fanno pensare che le nubi stiano fluendo verso il centro della galassia lungo la struttura a barra e questo potrebbe spiegare l'alto tasso di formazione di stelle nella regione. La distribuzione delle nubi molecolari è stata ricostruita sulla base dell'emissione radio a 2,6 millimetri delle molecole di monossido di carbonio da Kwok-Yung Lo e collaboratori del Cal Tech, con il nuovo interferometro per onde millimetriche ad alta risoluzione dell'Owens Valley Radio Observatory. L'immagine ottica a falsi colori è stata ottenuta con un dispositivo a scorrimento di carica applicato al telescopio da 60 pollici di Mount Wilson.

Un esame più approfondito dell'emissione del monossido di carbonio da M82 indica che la massima concentrazione di materia allo stato molecolare si trova al centro della galassia, vicino al picco di emissione nell'infrarosso lontano. La densità del gas molecolare supera quella dell'idrogeno atomico in tutto il disco ottico e per buona parte dei filamenti verticali. In effetti la massa totale del gas molecolare è pari a quasi metà della massa delle stelle, ossia circa tre volte il valore osservato nelle galassie a spirale normali. Sotto molti punti di vista le stranezze di M82 assomigliano molto alle caratteristiche che si dovrebbero osservare in una galassia giovane: una popolazione stellare con una forte percentuale di stelle di grande massa, a vita breve, una notevole abbondanza di gas interstellare non ancora condensato in stelle e un processo ancora in corso di caduta di gas intergalattico dalle regioni esterne della nube protogalattica. Siamo forse di fronte alla nascita - o, più probabilmente, a una seconda giovinezza - di un'intera galassia?

In alcune galassie, soprattutto in quelle che nella regione centrale hanno un tasso di formazione di stelle abbastanza ridotto, l'attività in quella regione potrebbe venire avviata e mantenuta dalla caduta di gas dalle regioni esterne del disco galattico. Un buon esempio di questo fenomeno è quello della galassia IC 342. Kwok-Yung Lo e collaboratori del Cal Tech hanno realizzato recentemente una mappa dell'emissione del monossido di carbonio nel nucleo di questa galassia con il nuovo interferometro per onde millimetriche ad alta risoluzione dell'Owens Valley Radio Observatory. I loro dati indicano per la prima volta con chiarezza che le nubi molecolari di IC 342 sono tutte concentrate in una struttura a forma di barra lunga circa 6000 anni luce. Le velocità rilevate delle nubi fanno pensare a un deciso moto di caduta verso il centro della galassia nel quale precedenti osservazioni nell'infrarosso avevano indicato un tasso di formazione di stelle superiore alla norma. Forse la barra funge da guida incanalando le nubi di gas verso la zona centrale.

A un livello di attività ancora superiore si colloca la galassia a spirale NGC 1068, appartenente a una classe di galassie che si distinguono per i loro nuclei visibili eccezionalmente luminosi e compatti, con forti righe di emissione dovute a gas ionizzato e ad alta velocità. La luminosità nell'infrarosso lontano raggiunge l'incredibile valore di 200 miliardi di unità solari e proviene tutta da una zona di 5000 anni luce di raggio intorno al centro. Questa galassia è anche ricchissima di gas molecolare, come si vede dalla intensa emis-

M82, una galassia molto particolare, qui è visibile ripresa a due diverse lunghezze d'onda. La fotografia in bianco e nero in alto è stata eseguita in luce rossa con il telescopio da 200 pollici di Monte Palomar e mostra un disco complessivamente amorfo con strisce di polvere scura che si stagliano contro la superficie e un fascio intricato di filamenti che si dipartono dal centro ortogonalmente al disco. L'immagine a colori riportata in basso, basata sull'emissione a 2,6 millimetri del monossido di carbonio, è stata ottenuta dagli autori. In questa immagine la luminosità è proporzionale all'intensità dell'emissione, mentre i colori rappresentano la velocità del gas relativamente al centro della galassia; il rosso corrisponde al gas che si allontana dall'osservatore, mentre il blu corrisponde al gas che si avvicina. I dati indicano che il gas del disco è in rotazione in un piano parallelo alla linea di vista; il gas dei filamenti, invece, sta cadendo verso il centro della galassia.

sione del monossido di carbonio. Ciononostante è chiaro che il livello di attività attuale non può essere mantenuto per più di qualche centinaio di milioni di anni se la luminosità proviene da stelle giovani e la riserva di materiale per la formazione di nuove stelle non supera quella visibile oggi nelle regioni centrali. D'altro canto è difficile capire come la riserva centrale possa essere rifornita mediante trasporto dalle regioni esterne del disco, in cui il gas si trova attualmente in un'orbita circolare. Di solito, quindi, si preferisce spiegare il comportamento di queste galassie attribuendolo a proliferazioni improvvise di stelle e ritenere sporadica l'attività delle regioni centrali.

Uno dei problemi più complessi relativamente ai modelli che si basano sulla proliferazione improvvisa di stelle è la natura dell'agente che regola il processo. Come ha inizio il fenomeno, perché si diffonde e in che modo l'«incendio» infine viene soffocato? Forse l'attività delle nubi molecolari distribuite in tutto il disco galattico interno viene innescata dall'azione dell'oggetto centrale: un buco nero, se ne esiste uno. Sembra che NGC 1068 presenti tutte e due queste forme di attività. Una sorgente di meno di 100 anni luce di diametro ha una luminosità infrarossa, ottica e X pari a quasi 100 miliardi di volte quella del Sole: presumibilmente questa energia ha origine in un oggetto centrale compatto. Parallelamente vi sono chiari segni del fatto che la luminosità nell'infrarosso lontano, di analoga potenza, provenga da un disco di nubi largo almeno alcune migliaia di anni luce. Forse la sorgente centrale stimola la formazione di stelle nelle nubi circostanti inviando attraverso la galassia onde d'urto esplosive. È anche possibile che questa regione rimanga inerte fino all'accumulo di una massa critica di materiale per la formazione di stelle; in questo caso l'avvio iniziale del processo di formazione in una piccola zona potrebbe non solo autosostentarsi, ma anche portare alla formazione di stelle di grande massa in una reazione a catena travolgente. Attualmente le osservazioni del gas molecolare riguardano un numero di galassie troppo esiguo per poter dire se esistono davvero galassie molto ricche di gas, ma con una scarsa attività di formazione stellare, come farebbe credere la seconda ipotesi che, peraltro, lascia aperto il problema della natura dell'oggetto centrale.

È stato ormai chiaramente stabilito che quasi tutte le stelle devono formarsi all'interno di nubi molecolari. Questa conclusione deriva dagli studi di regioni di formazione di stelle a noi vicine, come la nebulosa di Orione, e discende anche dalla forte correlazione scoperta nelle galassie esterne tra gas molecolare e radiazione proveniente da stelle giovani; l'altro principale componente gassoso di queste galassie, l'idrogeno atomico, non presenta infatti questa correlazione.

Nonostante le conoscenze ottenute negli ultimissimi anni gli astronomi si trovano ancora di fronte a problemi assai difficili da risolvere sulla natura dei bracci a spirale delle galassie. Che cos'è in realtà questo fenomeno che domina la morfologia visibile delle galassie? Un braccio è solo un mutamento di fase nel disco galattico, come le nubi bianche e gonfie dell'atmosfera terrestre, o è una struttura più concreta e materiale, un'onda di densità di materia che si propaga attraverso il disco galattico? Fino a oggi non esiste una risposta precisa, anche perché può darsi che la natura dei bracci cambi da una galassia all'altra, e persino da una regione all'altra della stessa galassia.

Gli studi osservativi del gas molecolare hanno dimostrato chiaramente che l'abbondanza di nubi molecolari può variare molto da una galassia all'altra e alle varie distanze dal centro della stessa galassia. Se l'esistenza di questo gas è una condizione necessaria per la formazione delle stelle, allora nelle galassie ricche di gas molecolare questa condizione sarà soddisfatta in tutto il disco. In questi casi l'andamento del fenomeno dovrebbe essere molto meno ordinato e coerente, perché le stelle si formeranno più o meno dovunque. D'altra parte, nelle galassie povere di gas molecolare, questa condizione può venir soddisfatta solo in certe zone. In queste galassie è molto più facile che si formi una grande struttura a spirale di formazione di stelle, con nubi molecolari solo lungo i bracci. È chiaro, quindi, che analizzando l'andamento su grande scala di formazione delle stelle bisogna esaminare separatamente le galassie ricche di gas molecolare, come la nostra, e quelle che ne contengono poco, come M31 (la galassia di Andromeda) o M81. Queste ultime due galassie vengono spesso citate come esempi di grandi strutture a spirale in accordo con la teoria delle onde di densità; entrambe sono anche molto povere di gas molecolare. La scarsa abbondanza di gas molecolare è in accordo con i risultati di Richard A. Linke e Anthony Stark dei Bell Laboratories secondo i quali l'emissione del monossido di carbonio in M31 è strettamente limitata ai bracci a spirale visibili.

Il problema fondamentale nello studio della struttura a spirale non è tanto quello di capire come si formi un braccio curvo, ma piuttosto come questa struttura si conservi per un lungo periodo di tempo e divenga simmetrica intorno al centro della galassia. È ben noto che qualunque perturbazione in moto radiale nel disco di una galassia a spirale sarà costretta ad assumere una forma a spirale semplicemente in conseguenza della rotazione della galassia. Di solito si vede che la velocità di rotazione di una galassia è abbastanza costante all'aumentare del raggio per quasi tutta l'estensione del disco, e questo comporta che la materia che orbita più lontano del centro rimanga indietro e quella che orbita più vicino al centro si sposti in avanti (rispetto a un punto di riferimento situato in una posizione intermedia). Quindi in queste galassie un fenomeno come un'esplosione o un'«ondata» di formazione di stelle che si propaghi da una nube all'altra formerebbe un braccio curvo in modo del tutto naturale. La difficoltà nasce quando si cerca di coordinare il fenomeno su tutta l'estensione del disco.

Oggi sembra possibile che buona parte della confusione che regna a questo proposito sia dovuta al tentativo di estendere la coerenza e la simmetria dell'andamento a spirale a tutte le galassie invece di accettare tranquillamente il fatto che vi sono alcune galassie, forse una maggioranza, nelle quali non c'è praticamente traccia di coerenza. In questo quadro le galassie con una configurazione a spirale davvero coerente sarebbero quelle come M31 e M81, che sono relativamente povere di nubi in cui si formano stelle, o quelle come M51, che ha una compagna abbastanza vicina da esercitare una forte azione mareale su tutto il disco galattico. Nelle altre galassie ricche di nubi la formazione di stelle risulterebbe più diffusa e meno organizzata.

Un fatto incontestabile è che i bracci, così come sono, sono zone di formazione di stelle di grande massa. Se poi a questo corrisponda un aumento del tasso di formazione di tutte le stelle nei bracci o solo una variazione del rapporto tra le stelle massicce e le altre non si sa. È legittimo sperare in una risposta a questa domanda entro i prossimi dieci anni, perché i telescopi per l'infrarosso a bordo dei satelliti raggiungeranno una risoluzione angolare sufficiente a separare con chiarezza i bracci e le regioni intermedie nelle galassie esterne.

Osservando la distribuzione su grande scala delle nubi molecolari nella nostra e nelle altre galassie abbiamo approfondito le conoscenze dei rapporti tra queste regioni fredde e dense e le proprietà globali delle galassie come morfologia e luminosità. Poiché le nubi molecolari sono i precursori della formazione di stelle, è oggi possibile ottenere moltissime informazioni su una fase cruciale del ciclo vitale delle stelle e cominciare a capire l'evoluzione delle galassie.

Si è supposto che quasi tutta la materia organica dell'universo (cioè i composti del carbonio più complessi del monossido di carbonio) si trovasse sulla superficie e nell'atmosfera dei corpi planetari. Le enormi masse delle nubi molecolari comportano invece che siano queste i serbatoi principali di materia organica. Inoltre, se mai il sistema solare si trovasse ad attraversare una di queste nubi, l'assorbimento della luce sarebbe così forte da far scomparire alla vista tutte le stelle tranne le poche più vicine, e se il Sole viaggiasse a una velocità stellare tipica intorno ai 34 000 chilometri all'ora, sarebbero necessari più di due milioni di anni per uscire dalle tenebre. Data l'abbondanza delle nubi nella nostra galassia, un avvenimento del genere dovrebbe verificarsi circa una volta ogni miliardo di anni, cioè circa cinque volte nel corso della storia della Terra. Se l'uomo si fosse evoluto durante uno di questi episodi, la sua visione dell'universo e la sua concezione filosofica sarebbero state radicalmente diverse.

Incudini di diamante per alte pressioni

Un ingegnoso dispositivo consente di riprodurre la pressione esistente nelle vicinanze del nucleo della Terra facendo girare una vite; le sue applicazioni hanno rivoluzionato lo studio delle fasi ad alta pressione

di A. Jayaraman

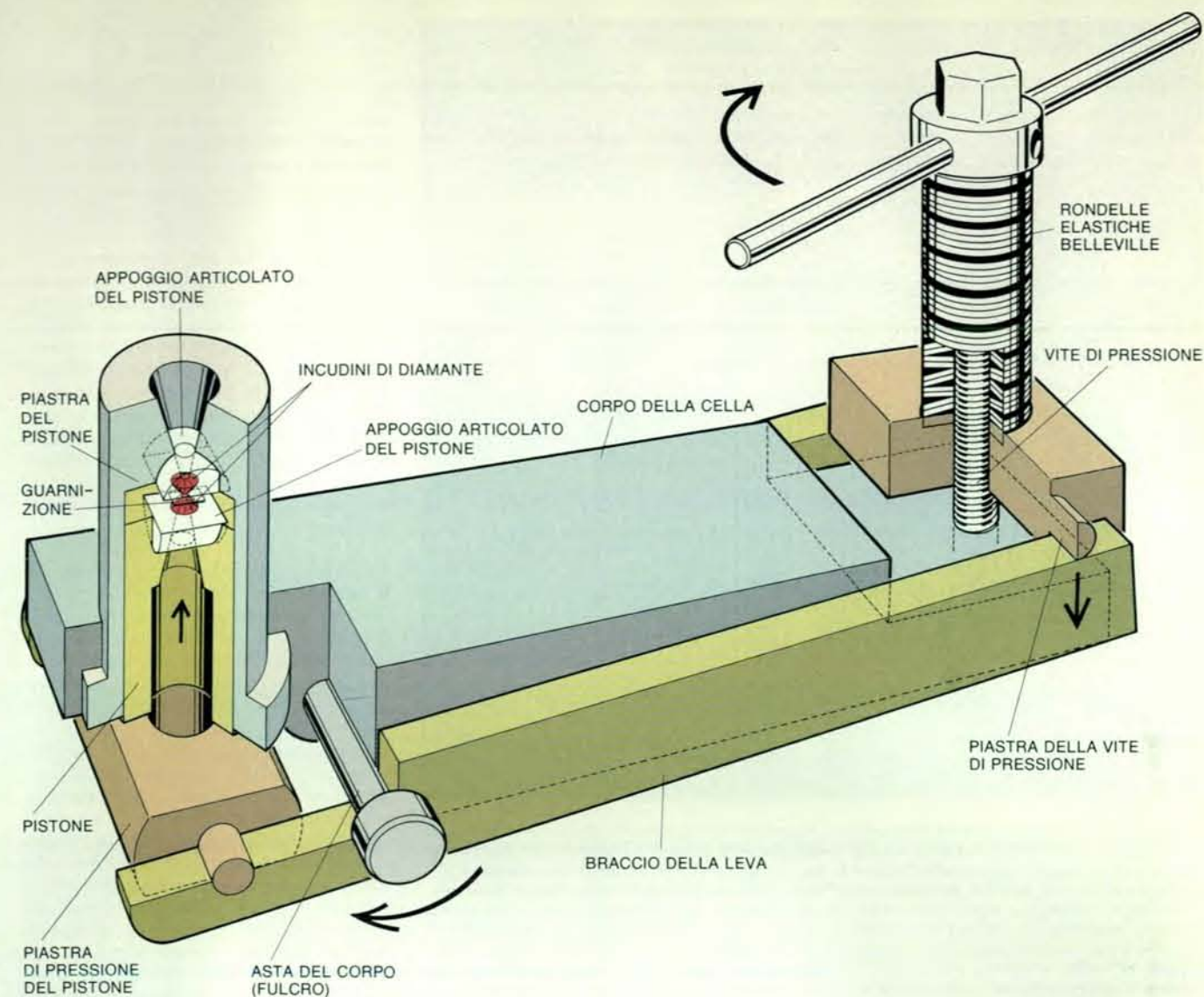
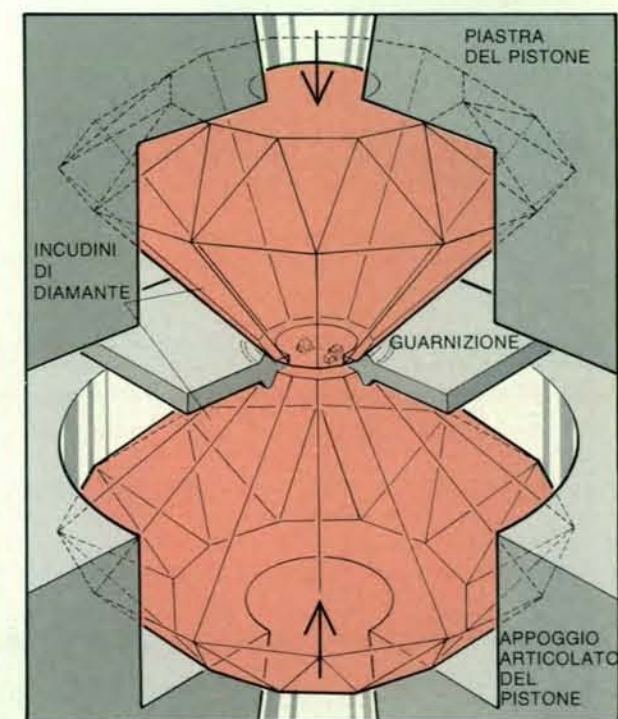
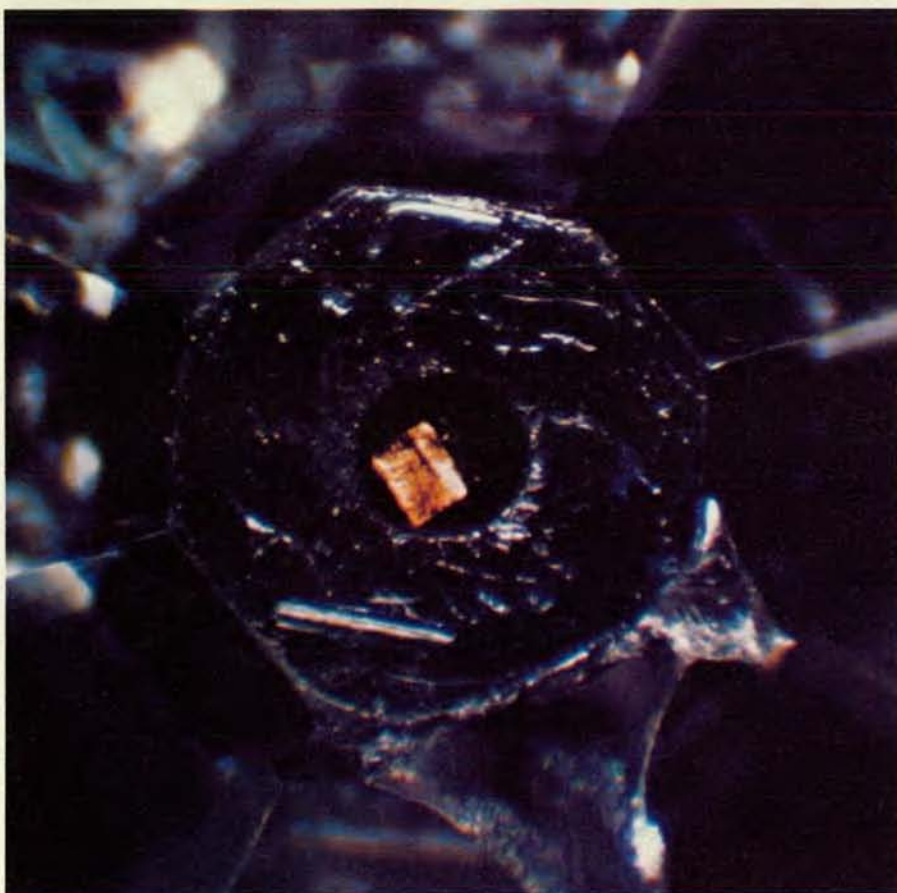
Nel retro del mio laboratorio presso gli AT&T Bell Laboratories c'è un'enorme apparecchiatura idraulica, alta due metri, del peso di parecchie tonnellate e sostenuta da quattro robusti pilastri di acciaio. Una sostanza solida posta nell'apparecchiatura può essere sottoposta a una pressione 10 000 volte superiore a quella dell'atmosfera terrestre al livello del mare, oltre 100 volte superiore alla pressione sul fondo delle più profonde fosse oceaniche. A tali pressioni le forze applicate alla materia solida sono tanto intense che il reticolo degli atomi che formano il solido si può deformare e gli elettroni vaganti nel solido possono assumere stati energetici insoliti. Queste variazioni microscopiche possono alterare profondamente le proprietà macroscopiche di una sostanza: la sua densità, il suo colore, la sua opacità, la sua duttilità, la sua conducibilità elettrica e la sua sensibilità a un campo magnetico. Inoltre, l'esplorazione di tali variazioni in laboratorio non è affatto una semplice esercitazione di fisica. La maggior parte della materia dell'universo, e quindi la maggior parte della materia che costituisce la Terra, è confinata a pressioni ben superiori a 100 000 atmosfere. Risulta che sarebbe necessaria una «pressa» molto più grande di quella del retro del mio laboratorio per raggiungere le pressioni normali delle regioni più profonde del pianeta.

Mi si consenta di descrivere il resto del mio laboratorio. Le mie ricerche riguardano quasi esclusivamente le proprietà della materia a pressioni enormi, ma la maggior parte dello spazio del laboratorio è oggi destinato a strumenti che hanno ruoli secondari nella ricerca: il laser, il microscopio, lo spettroscopio e l'apparecchiatura automatica che può registrare i dati spettroscopici. In effetti, sarebbe facile non notare il nuovo dispositivo che rende possibile il resto della ricerca, dal momento che lo si può agevolmente tenere nel palmo della mano. Nonostante le sue dimensioni, questo dispositivo può creare regolarmente pressioni da 10 a 15 volte maggiori della più elevata pressione ottenibile con la pressa idraulica. La massima pressione ottenuta finora con il dispositivo è di 1,7 milioni di atmosfere, equivalente alla pressione esistente 2900 chilometri al di sotto della superficie terrestre.

Al centro del dispositivo vi sono due diamanti intagliati purissimi, ciascuno di peso compreso tra un terzo di carato e mezzo carato. Le punte dei diamanti sono lavorate a forma di piccole superfici piane affacciate una all'altra. I due diamanti sono montati all'interno di un meccanismo, di acciaio temprato e lavorato fino ad alte tolleranze, il quale fornisce una spinta lungo un asse perpendicolare alle facce dei diamanti. Vi sono almeno cinque progetti per il meccanismo di spinta, ma in linea di principio si tratta sempre di poco più di un elaborato schiaccianoci. Un piccolo campione del materiale da studiare ad alta pressione viene posto tra le facce dei diamanti; si può poi applicare manualmente una forza, che viene moltiplicata da 500 a 1000 volte grazie al meccanismo ed è quindi trasmessa al campione inserito tra i diamanti. La superficie sulla quale la forza viene concentrata è estremamente piccola e di conseguenza la pressione agente sul campione, ovvero la forza per unità di superficie, può essere enorme. Ogni dispositivo del genere, nel quale due facce di diamante contrapposte applicano una pressione a un campione, viene detto cella a incudini di diamante.

Il semplice funzionamento della cella a incudini di diamante è il risultato di una dura battaglia ventennale nell'evoluzione dei progetti. È relativamente facile produrre alte pressioni dinamiche per pochi milionesimi di secondo con un'onda d'urto; la produzione di alte pressioni statiche invece è problematica tanto quanto lo studio delle proprietà dei materiali sottoposti a tali pressioni. Un problema fondamentale è il contenimento del campione del materiale in esame: le pareti del recipiente sono sottoposte alle stesse pressioni del campione e si deve quindi tenere presente la possibilità di deformazione o di rottura del recipiente. Inoltre, dato che la pressione sul campione deve essere idrostatica, cioè uguale in tutte le direzioni, il campione all'interno della cella ad alta pressione può essere circondato da un materiale molle o fluido. Se il campione e il materiale idrostatico devono essere caricati convenientemente in una cella riutilizzabile, deve esistere un sistema di guarnizioni in grado di sigillare la cella caricata contro il pericolo di eventuali fughe. Infine, le proprietà del campione si devono determinare mentre la cella è sotto pressione; se si devono studiare gli effetti di variabili come la temperatura alle alte pressioni, anch'essi devono essere controllati e misurati senza ridurre la pressione all'interno della cella.

P. W. Bridgman della Harvard University diede il via allo studio della materia alle alte pressioni quasi da solo per più di mezzo secolo fino alla morte, nel 1961. Bridgman non solo studiò un notevole numero di elementi e composti ad alta pressione, ma inventò anche tutte le tecniche con le quali lavorava. Introdusse un sistema di guarnizioni per confinare un campione tra due incudini contrapposte lungo una sola direzione e fece uso di incudini di carburo di tungsteno, un materiale estremamente duro a elevata intensità di compressione. Per mantenere la pressione idrostatica circondava il campione nella camera ad alta pressione con cloruro di argento, un materiale solido molle che può trasmettere uniformemente la pressione sulla superficie del campione.

Nel 1905, quando era ancora studente, Bridgman scoprì con grande sorpresa che la camera ad alta pressione tra due incudini poteva essere sigillata con un rivestimento di materiale molle. La pressione sul materiale di rivestimento era sempre maggiore di quella all'interno della camera e quindi il rivestimento funzionava come guarnizione che non avrebbe consentito perdite per quanto alta fosse stata la pressione. Tali

Il meccanismo di spinta della cella a incudini di diamante messa a punto presso il National Bureau of Standards è un semplice dispositivo che sta nel palmo di una mano, costruito con alte tolleranze da un pezzo di acciaio temprato. L'apparecchiatura illustrata qui sopra è una versione leggermente modificata di una analoga messa a punto presso il Geophysical Laboratory della Carnegie Institution di Washington; è lunga circa 20 centimetri e pesa circa tre chilogrammi. Il suo funzionamento è in linea di principio simile a quello di uno schiaccianoci. La spinta viene prodotta facendo ruotare in senso orario la testa di una vite. La spinta viene poi trasmessa dal braccio di una leva a un pistone mobile la cui testa è rientrata per sostenere la superficie arrotondata di un sistema di appoggio articolato semicilindrico. Un diamante viene fissato sulla superficie piana del semicilindro, mentre la punta del diamante viene resa piatta. La faccia piana di questo diamante, che è il diamante inferiore mostrato nell'ingrandimento a destra, viene spinta contro la faccia piana del diamante superiore in modo da produrre una pressione nella cella ad alta pressione. Anche il diamante superiore è montato su un sistema di appoggio articolato semicilindrico, disposto perpendicolarmente al semicilindro inferiore; il parallelismo delle due facce dei diamanti viene quindi garantito muovendo i due semicilindri fino a quando le frange di interferenza ottica generate dalle due facce si riducono a una figura di un grigio uniforme. La cella non è altro che un forellino praticato in una lamina di acciaio montata tra le due facce. Nella cella vengono sistemati un campione del materiale da studiare ad alta pressione, un rubino per la calibrazione della pressione e una goccia di fluido per trasmettere uniformemente la pressione in tutte le direzioni. Applicando la pressione l'acciaio subisce un'estrusione attorno alle facce dei diamanti e funge quindi da guarnizione per il contenimento del materiale della cella. La pila di rondelle elastiche tipo Belleville agisce come una molla per caricare la vite di regolazione della pressione. Ogni rondella elastica ha la stessa forma di una rondella ordinaria, ma presenta una sporgenza lungo il suo asse perpendicolare; le rondelle si possono sovrapporre in modo che le sporgenze siano rivolte nella stessa direzione o in direzione opposta lungo la vite. Spostando le rondelle elastiche si può modificare la rigidità della molla e si può quindi variare l'entità della spinta per ogni giro di vite. Il dispositivo mostrato produce pressioni di 1,7 milioni di atmosfere.

Un cambiamento della fase solida in un cristallo di monosolfuro di samario (SmS) dà origine a uno spettacolare cambiamento di colore e a una variazione delle proprietà elettriche del cristallo. In questa microfotografia eseguita dall'autore la fase dorata ad alta pressione della sostanza si può osservare attraverso uno dei diamanti lungo l'asse di spinta della cella a incudini di diamante. Attraverso il diamante sono visibili anche le sfaccettature del diamante stesso, leggermente sfuocate a causa della limitata profondità di campo della microfotografia. Il monosolfuro di samario è un semiconduttore nero opaco a pressione atmosferica; la banda di valenza, in altri termini la banda energetica più popolata da elettroni nel materiale, è completamente piena, mentre la banda energetica immediatamente successiva è vuota. Tra le due bande esiste anche un netto livello energetico popolato da elettroni. A una pressione di 7000 atmosfere il colore nero diventa dorato perché gli elettroni del livello netto vengono schiacciati nella banda energetica vuota e il materiale comincia a condurre la corrente elettrica, comportandosi come un metallo. Riducendo la pressione, la fase dorata del cristallo si ritrasforma nuovamente nella fase nera.

fondamentali perfezionamenti tecnici permisero a Bridgman di studiare le proprietà della materia fino a pressioni di 100 000 atmosfere.

Le incudini di Bridgman hanno un limite nella comprimibilità del carburo di tungsteno e nell'opacità del materiale a tutte le radiazioni. Può perciò apparire sorprendente che non venisse impiegato fin dall'inizio il diamante come materiale per l'incudine. Ovviamente, ai tempi di Bridgman era risaputo che il diamante è trasparente ai raggi X ed è tuttora la sostanza più dura che si conosca. Ciononostante, solo a partire dal 1959 il diamante venne impiegato come materiale per incudini negli studi delle alte pressioni da Charles E. Weir, Ellis R. Lippincott, Alvin Van Valkenburg, Jr., ed Elmer N. Bunting del National Bureau of Standards e indipendentemente da John C. Jamieson, Andrew W. Lawson e Norman D. Nachtrieb dell'Università di Chicago. Da allora i ricercatori del Bureau of Standards e di altri laboratori hanno progressivamente perfezionato il dispositivo originale trasformandolo in un finissimo strumento per lo studio quantitativo delle alte pressioni.

Come viene caricato un campione nella cella a incudini di diamante e come viene compresso a una pressione prefissata? Nel mio laboratorio ho la fortuna di avere l'assistenza di Ralph G. Maines, che ha avuto un addestramento da orologiaio ed è dotato di una mano abilissima con strumenti di alta precisione da manipolarsi al microscopio. Il primo passo è la preparazione della cella ad alta pressione, che è semplicemente un foro praticato in una lamina metallica. Quando la lamina metallica viene montata tra le facce dei due diamanti e il foro viene allineato con l'asse di spinta, il materiale nel foro può essere sottoposto a un'alta pressione. Il metallo che circonda il foro viene spinto lateralmente alle facce del diamante quando si applica la pressione, e quindi il metallo si comporta come una guarnizione che sigilla lo spazio attorno al campione *(si veda l'illustrazione a pagina 57)*.

Per poter centrare il foro sull'asse di spinta, il metallo viene intagliato dai diamanti, e il foro, di soli 200 micrometri di diametro, viene praticato il più vicino possibile al centro dell'intaglio. La cella ad alta pressione che ne risulta è un cilindro del diametro di 200 micrometri con lo stesso spessore della lamina metallica compressa: circa 100 micrometri. La guarnizione è sistemata sulla faccia del diamante inferiore con la stessa orientazione che aveva al momento dell'intaglio. Vengono quindi sistemati nel foro il materiale campione e un piccolo frammento di rubino per la calibrazione della pressione. Infine, per poter sostenere la pressione idrostatica, il foro viene riempito con una goccia di fluido da una siringa e poi immediatamente sigillato dalle facce dei diamanti prima che il fluido possa evaporare.

Il rubino posto nella cella con il campione consente di determinare la pressione nella cella stessa con un elevato grado di precisione. In effetti, in mancanza del rubino le piccole dimensioni della cella ad alta pressione renderebbero molto difficile la calibrazione della pressione, e il valore scientifico dello strumento sarebbe molto inferiore. Quando viene eccitato dalla luce, un rubino emette per fluorescenza un intenso rosso scuro. L'emissione per fluorescenza può essere risolta con uno spettrografo in due picchi le cui lunghezze d'onda sono note con precisione alla pressione atmosferica. Quando si aumenta la pressione sul rubino, entrambi i picchi di fluorescenza si spostano verso una lunghezza d'onda maggiore. Lo spostamento è stato calibrato per pressioni note da misure indipendenti, quindi lo spostamento dei picchi misurato nello spettro è una misura indiretta della pressione agente sul rubino. Lo spostamento spettrale dei picchi è quasi esattamente proporzionale alla pressione, per pressioni fino a 300 000 atmosfere; per pressioni superiori lo spostamento spettrale è lievemente minore a parità di incremento di pressione, ma si può ancora valutare la pressione dalla fluorescenza.

Sono molti i materiali comunemente impiegati come mezzo idrostatico nella cella. Il materiale più comodo da caricare in una cella è un miscuglio di quattro parti di metanolo e una parte di etanolo, ma a pressioni superiori a 104 000 atmosfere il miscuglio congela a temperatura ambiente e la sua azione non è più rigorosamente idrostatica. Si possono mantenere pressioni idrostatiche fino a circa 300 000 atmosfere immergendo la cella in azoto liquido e intrappolando poi nella cella un gas nobile quale lo xeno. Lo xeno congela a circa 112 gradi sotto lo zero, ma è un mezzo eccellente per pressioni idrostatiche nella sua forma solida e non reagisce chimicamente con il campione. A pressioni superiori sembra che l'elio e l'idrogeno solidi siano idrostatici fino ad almeno 700 000 atmosfere, ma richiedono o che la cella sia raffreddata fino a una tem-

peratura solo di pochi gradi superiore allo zero assoluto o che esista un mezzo per fornire l'idrogeno o l'elio alla cella ad alta pressione.

Una volta che la cella è stata caricata e pressurizzata, il diamante agisce come una finestra sul campione ed eventuali variazioni di colore o di luminosità del campione con la pressione si possono osservare direttamente al microscopio. Nel 1970 venne scoperta, per il composto monosolfuro di samario (SmS), una spettacolare transizione visiva. A pressione atmosferica la sostanza si presenta come un opaco semiconduttore nero: un materiale la cui conducibilità elettrica è intermedia tra quella di un metallo e di un isolante. Però, all'aumentare della pressione fino a 7000 atmosfere, il monosolfuro di samario comincia a brillare come l'oro e il materiale conduce una corrente elettrica con la bassa resistenza tipica di un metallo. Se, successivamente, la pressione viene abbassata, ricompare il semiconduttore nero opaco; i colori del campione osservato al microscopio passano alternativamente dal nero al dorato girando semplicemente una vite che agisce sulla cella a incudini di diamante.

Le variazioni di colore e della conducibilità elettrica del monosolfuro di samario sono il risultato di una transizione di fase indotta dalla pressione sul materiale. Le fasi più comuni di una sostanza sono quella solida, liquida e gassosa, ma nella maggior parte delle sostanze vi sono anche molte fasi solide distinte. Una fase solida stabile (più in generale, una qualsiasi fase stabile di un sistema) si dice in equilibrio termodinamico con l'ambiente circostante. Dopo una prolungata esposizione alla temperatura e alla pressione dell'ambiente, l'energia totale (cioè la somma dell'energia cinetica e dell'energia potenziale) immagazzinata da una sostanza deve raggiungere un valore minimo. L'energia totale di un solido può diminuire grazie a una variazione di configurazione degli atomi che formano il solido o a una variazione della struttura energetica degli elettroni del solido. Pertanto, quando il volume di un solido viene ridotto per compressione, è possibile che la sua struttura atomica o la sua struttura elettronica debbano cambiare, perché raggiunga l'equilibrio energetico con il nuovo ambiente. Tali variazioni possono modificare sensibilmente le proprietà fisiche del solido. Per lo studioso della fisica delle alte pressioni la scoperta di una transizione di fase indotta dalla pressione è un'esperienza profondamente gratificante.

Una delle previsioni più stimolanti della teoria dei solidi è che a pressioni comprese tra un milione e parecchie centinaia di milioni di atmosfere molte sostanze, gassose a temperatura ambiente e a pressione atmosferica, diventano solidi metallici. Per esempio, si ritiene che l'idrogeno diventi metallico a circa due milioni di atmosfere e che l'elio diventi metallico in prossimità dei 110 milioni di atmosfere. Pressioni statiche così elevate non sono state raggiunte, ma la cella a incudini di diamante consente di verificare la teoria sulla quale si basano queste previsioni in condizioni sperimentali più accessibili.

La struttura a bande di energia degli elettroni determina se un solido è o non è metallico. Nell'atomo isolato gli elettroni occupano livelli energetici discreti, ma quando molti atomi si aggregano in un solido, questi livelli si allargano in bande di energia. L'allargamento è un effetto quantomeccanico: si manifesta perché due elettroni che occupano lo stesso livello energetico in due atomi ampiamente separati devono assumere livelli energetici leggermente diversi quando vengono confinati in un piccolo volume. Le bande energetiche elettroniche di un solido sono perciò formate da molti, forse 10^{23}, livelli energetici discreti, ma i livelli sono raggruppati insieme così fittamente da non potersi risolvere. Le bande energetiche adiacenti di un solido possono sovrapporsi oppure essere separate da un intervallo, che rappresenta un insieme di valori energetici «proibiti» per gli elettroni del solido.

La trasparenza del diamante permette di eseguire precise misurazioni ottiche sulle proprietà di un materiale campione confinato ad alte pressioni nella cella a incudini di diamante. La complessa apparecchiatura sperimentale nel diagramma illustra alcune delle misurazioni che si possono eseguire. La pressione si determina focalizzando il fascio di un laser a elio-cadmio *(in blu)* su un rubino nella cella, provocando così la fluorescenza del rubino *(in rosso)*. Uno spettrometro misura lo spostamento in lunghezza d'onda della fluorescenza del rubino con la pressione, e tale spostamento fornisce il valore della pressione nella cella. Il fascio di un laser a ioni argo *(in verde)* può essere focalizzato sul campione che risponde emettendo la luce chiamata diffusione Raman *(in giallo)* di frequenza inferiore a quella della luce verde incidente. La sovrapposizione nello spazio delle emissioni rossa e gialla è in arancione. L'analisi della luce di diffusione Raman con il doppio monocromatore (uno spettrometro ad altissima risoluzione) consente di determinare la frequenza vibrazionale dei legami interatomici e di misurare quindi la forza dei legami a varie pressioni. Infine, il campione può essere riscaldato fino a 2000 gradi centigradi con un fascio *(in viola)* prodotto da un terzo laser, il laser a neodimio-YAG (ittrio-alluminio-granato). La temperatura viene misurata con un pirometro ottico, un dispositivo nel quale il colore incandescente del campione riscaldato viene confrontato a vista con il colore del filamento di una lampadina a incandescenza di temperatura nota. Si può così studiare il campione alle elevate temperature e pressioni esistenti alle profondità della Terra.

La calibrazione della pressione nella cella ad alta pressione si basa sulla variazione della lunghezza d'onda di fluorescenza del rubino con la pressione. Un aumento di pressione provoca un aumento proporzionale della lunghezza d'onda per pressioni fino a 300 000 atmosfere *(curva continua in colore)*. Per pressioni più elevate si riscontra una modesta deviazione dalla proporzionalità *(curva nera tratteggiata)*. La scala è stata messa a punto da J. Dean Barnett, Stanley Block e Gasper J. Piermarini del Bureau of Standards e da Ho-kwang Mao e Peter M. Bell del Geophysical Laboratory.

Se gli elettroni più energetici di un solido riempiono completamente una delle bande di energia, e se c'è un intervallo proibito tra la banda piena e quella di energia immediatamente più elevata, il solido è un isolante elettrico. Se invece la banda di energia più elevata è piena solo parzialmente o se è vuota e ciononostante si sovrappone a una banda completamente occupata, il solido è un metallo e conduce facilmente la corrente elettrica. La previsione teorica fondamentale è che l'intervallo proibito che porta all'isolamento elettrico in un solido possa essere diminuito alle alte pressioni per finire con lo scomparire del tutto. Una volta che la soglia energetica della banda di energia vuota ha attraversato il confine della banda piena si raggiunge una condizione sufficiente per la conduzione metallica e l'isolante dovrebbe diventare metallico. Calcoli del genere eseguiti per l'idrogeno e l'elio indicano che questo evento dovrebbe accadere in altri materiali a pressioni raggiungibili con la cella a incudini di diamante. Per esempio, nello xeno la transizione metallica è prevista a circa 1,3 milioni di atmosfere, mentre nello iodio molecolare è prevista a 200 000 atmosfere.

Lo xeno solido è un isolante elettrico. L'ampiezza in energia dell'intervallo proibito associata alle sue proprietà di isolante è molto alta: nove elettronvolt a pressione atmosferica. L'ampio intervallo proibito inoltre rende lo xeno solido trasparente alla luce. I fotoni luminosi vengono assorbiti da un isolante solo se sono dotati di energia sufficiente a far superare a un elettrone l'intervallo proibito; l'assorbimento di fotoni a energia inferiore richiederebbe che un elettrone occupi l'intervallo proibito, il che è impossibile. Tuttavia l'energia dei fotoni più energetici della luce visibile è di circa tre elettronvolt solamente, pari a un terzo dell'intervallo proibito dello xeno. Il restringimento dell'intervallo proibito, un prerequisito per la transizione da isolante a metallo, dovrebbe perciò essere visibile nella cella a incudini di diamante. Lo xeno solido, trasparente a pressione atmosferica, dovrebbe diventare opaco alla luce al crescere della pressione.

Recenti misure spettroscopiche dell'assorbimento di luce da parte di xeno solido indicano che l'energia dell'intervallo proibito diminuisce a circa quattro elettronvolt a una pressione di 600 000 atmosfere. Questa diminuzione dell'energia dell'intervallo proibito con la pressione è in accordo con le previsioni sul senso della variazione, ma lo xeno a 600 000 atmosfere resta un isolante trasparente. Nessuno ha mai visto lo xeno solido diventare opaco alla luce. Nel 1979, però, fu misurata la resistività elettrica di una sottile pellicola di xeno comprimendo la pellicola tra una piccola punta sferica di diamante e una faccia di diamante piana. La pressione sullo xeno in tale dispositivo non può essere calibrata con la stessa precisione che nella cella a incudini di diamante ma, per una pressione stimata superiore a 1,3 milioni di atmosfere, la resistività dello xeno risultò molto bassa. È perciò probabile che nello xeno sia stata osservata la transizione alla fase metallica, ma sono necessarie ulteriori conferme sperimentali.

Per lo iodio la transizione dalla fase di isolante a quella metallica è invece accertata senza alcun dubbio. Alla pressione atmosferica lo iodio è un cristallo elettricamente isolante formato da molecole di iodio biatomiche. Vent'anni fa Harry G. Drickamer e i suoi collaboratori dell'Università dell'Illinois studiarono la resistenza elettrica dello iodio sotto pressione. Essi trovarono una grande e continua diminuzione della resistenza all'aumentare della pressione fino a 200 000 atmosfere. Sulla base di questo risultato avanzarono l'ipotesi che lo iodio potesse diventare metallico sotto pressione, ma non furono in grado di stabilire se il cristallo di iodio metallico è formato da atomi o da molecole biatomiche.

Solo negli ultimi anni gli studi delle alte pressioni con la cella a incudini di diamante hanno risolto la questione. Dal momento che il diamante è trasparente ai raggi X, si può studiare la struttura atomica dei materiali sotto pressione nella cella con il metodo della diffrazione dei raggi X. La figura dei raggi X riflessi dal reticolo atomico di un materiale consente di dedurre la struttura e la configurazione del reticolo. Oggi è chiaro, grazie a studi di figure di diffrazione dei raggi X, che lo iodio diventa gradualmente metallico in forma molecolare cristallina perché cominciano a sovrapporsi una banda energetica elettronica piena e una banda vuota. Tuttavia, a una pressione di circa 210 000 atmosfere il cristallo molecolare viene bruscamente trasformato in uno monoatomico e la struttura delle bande elettroniche cambia interamente. Il cristallo resta metallico nella sua fase monoatomica, ma l'elevata conducibilità elettrica di tale fase è l'effetto di una banda parzialmente piena e non l'effetto della sovrapposizione di due bande.

I tre stadi della trasformazione in fase metallica dello iodio sono diventati importanti per gli studi sperimentali del comportamento dell'idrogeno solido. Vi è un notevole interesse nella produzione di idrogeno metallico, perché è previsto su basi teoriche che l'idrogeno metallico debba essere superconduttore a temperature relativamente alte; in altri termini, esso dovrebbe condurre corrente elettrica senza resistenza a temperature significativamente superiori a quelle necessarie per far diventare superconduttori altri materiali. Inoltre, si ritiene che l'interno di pianeti giganti quali Giove, Saturno, Urano e Nettuno sia formato in gran parte da idrogeno. Conoscere le proprietà dell'idrogeno ad alta pressione è perciò importante per lo studio dei pianeti.

A temperatura ambiente l'idrogeno molecolare condensa e forma un limpido solido isolante e di alta comprimibilità a una pressione di quasi 57 000 atmosfere. Il solido è un cristallo molecolare perché la sua unità fondamentale è la molecola di

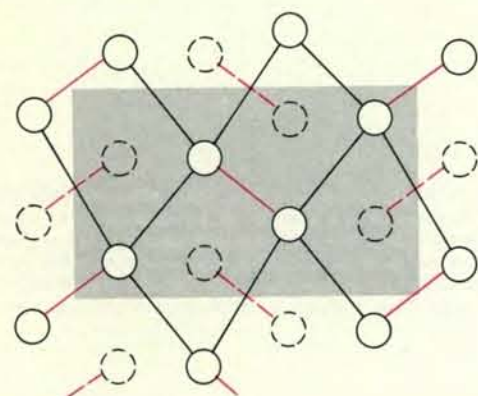
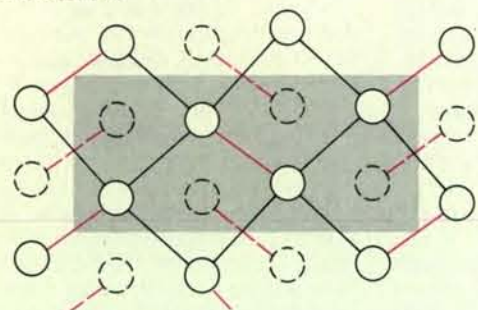
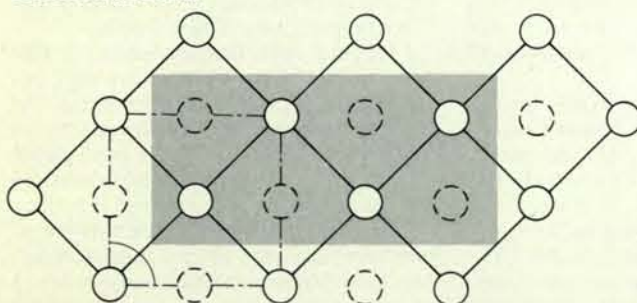
idrogeno biatomica (H_2). Una delle domande più interessanti negli studi attuali sull'idrogeno solido viene sollevata dall'analogia con la transizione a metallo dello iodio molecolare solido: l'idrogeno si comporta come lo iodio e quando diventa metallico si dissocia in un reticolo monoatomico o rimane nella forma molecolare per diventare metallico attraverso la sovrapposizione di due bande energetiche?

Due gruppi di ricercatori hanno iniziato lo studio dell'idrogeno e del suo isotopo deuterio alle alte pressioni. Presso il Geophysical Laboratory della Carnegie Institution di Washington è stata determinata a varie pressioni mediante la spettroscopia Raman la forza del legame nell'idrogeno molecolare e nel deuterio molecolare (idrogeno pesante, o idrogeno 2). Questa tecnica spettroscopica, che ha preso il nome dal fisico indiano C. V. Raman, si basa sulla misura della frequenza della luce diffusa da un campione sottoposto ad alte pressioni; la misura non si sarebbe quindi potuta effettuare se non fosse stato per la trasparenza dei diamanti che confinano il campione nella cella a incudini di diamante. La forza di un legame molecolare può essere paragonata alla rigidità di una molla a spire e la rigidità di una molla origina una frequenza caratteristica alla quale la molla vibra allungandosi e comprimendosi. Si può perciò determinare la forza del legame molecolare tra due atomi misurando la frequenza vibrazionale del legame.

Per poter effettuare la misura, un intenso fascio di luce blu ad alta frequenza emesso da un laser viene focalizzato sul campione pressurizzato nella cella a incudini di diamante. La luce blu interagisce con lo stato vibrazionale delle molecole del campione e appare come luce verde diffusa, che ha una frequenza inferiore a quella della luce blu. La differenza tra la frequenza della luce blu e quella della luce verde è esattamente uguale alla frequenza vibrazionale del legame molecolare. Quindi, misurando la frequenza della luce diffusa con uno spettrometro, si può ricavare la frequenza vibrazionale del legame. Lo spostamento della luce incidente a una frequenza inferiore è noto come effetto Raman.

I ricercatori del Geophysical Laboratory hanno scoperto che a 300 000 atmosfere la frequenza vibrazionale del legame interatomico nell'idrogeno molecolare comincia a decrescere; nel deuterio molecolare la frequenza comincia a diminuire a una pressione superiore a 400 000 atmosfere. Questi risultati mostrano che la forza dei legami diminuisce a tali pressioni, e i legami continuano a indebolirsi almeno fino a 700 000 atmosfere. Sembra quindi probabile che, sottoposto a una pressione ancora superiore, il cristallo di idrogeno molecolare finisca per diventare monoatomico; la trasformazione dell'idrogeno in metallo può dunque seguire le stesse fasi della trasformazione dello iodio.

Un secondo gruppo all'Università di Amsterdam ha studiato la variazione di volume dell'idrogeno molecolare solido all'aumentare della pressione. Fino a poco tempo fa la comprimibilità dell'idrogeno era stata determinata sperimentalmente soltanto per pressioni fino a 20 000 atmosfere, ma Isaac F. Silvera e collaboratori hanno ampliato le conoscenze fino a pressioni di 370 000 atmosfere. Il loro metodo è semplice e diretto. Il rubino per la calibrazione della pressione è posto nel foro della guarnizione, mentre il resto del foro viene riempito di idrogeno. Per ogni aumento di pressione si misura con un microscopio il diametro del foro e si determina lo spessore del foro dalle figure di interferenza della luce os-

1 ATMOSFERA

200 000 ATMOSFERE

300 000 ATMOSFERE

La struttura dello iodio è qui indicata a tre diverse pressioni. A sinistra è mostrata la configurazione degli atomi di iodio nel solido cristallino, e a destra è riportata la corrispondente struttura energetica elettronica. A pressione atmosferica l'unità fondamentale del cristallo è la molecola biatomica (I_2); il solido è un semiconduttore perché esiste un intervallo proibito tra la banda di valenza, che è piena di elettroni *(in colore)*, e la banda di conduzione, che è vuota. Aumentando la pressione fino a 200 000 atmosfere le molecole assumono una configurazione più ordinata e compatta e l'intervallo proibito si restringe finché le bande non si sovrappongono. Il cristallo molecolare diventa un metallo elettricamente conduttore. A circa 210 000 atmosfere avviene una brusca transizione a una fase metallica atomica; a 300 000 atmosfere gli atomi di questa fase formano un cristallo di elevata regolarità e la banda di valenza si riempie parzialmente, come le bande di energia più energetiche di un comune metallo. Gli atomi e i legami nel piano della figura sono mostrati come cerchi e linee continui, quelli non contenuti nel piano della figura come cerchi e linee tratteggiati. I legami intermolecolari sono in colore; gli altri legami tra gli atomi sono in nero. Le strutture dello iodio alle alte pressioni sono state determinate da K. S. Takemura, S. Minomura, O. Shimomura e Y. Fujii dell'Istituto per la fisica dello stato solido di Tokyo.

La struttura della parte superiore del mantello terrestre, che si estende dalla base della crosta terrestre fino a una profondità di circa 670 chilometri, è stata studiata sperimentalmente accoppiando una cella a incudini di diamante a un laser per riscaldare il materiale della cella. Sono riportate qui, in funzione della profondità e della pressione corrispondente all'interno della Terra, le transizioni di fase della forsterite, un componente dell'olivina, e della clinoenstatite, un membro del gruppo dei pirosseni. I nomi spinello, perovskite e ilmenite indicano solo strutture cristalline e si riferiscono a minerali con la stessa composizione chimica della forsterite e della clinoenstatite e la stessa forma cristallografica dei minerali della superficie terrestre che solitamente hanno tali nomi. La sequenza delle transizioni di fase indicata si basa su ricerche di Ling-gun Liu dell'Australian National University. Benché la forsterite e la clinoenstatite siano due tra i più abbondanti minerali del mantello, la presenza di ossidi di ferro, alluminio e calcio può far spostare le soglie di transizione e dare origine a ulteriori fasi ad alta pressione. La composizione del mantello inferiore, da una profondità di circa 670 chilometri fino al confine mantello-nucleo a 2900 chilometri, non è nota, ma si assume spesso che sia la stessa del mantello superiore.

servabile attraverso il foro stesso. Dal momento che la pressione si determina dallo spostamento dello spettro di fluorescenza del rubino e la variazione di volume del rubino si ottiene da informazioni indipendenti, si può rappresentare graficamente il volume di idrogeno che riempie la restante parte della cella in funzione della pressione.

I risultati di Silvera sono confermati da una misurazione completamente diversa della variazione di volume con la pressione eseguita al Geophysical Laboratory. A qualsiasi temperatura superiore allo zero assoluto una parte dell'energia termica di un solido viene trasportata da onde sonore, ovvero onde periodiche di compressione e rarefazione della materia che si propagano nel solido in tutte le direzioni. Le velocità di queste onde sonore si possono misurare con il sensibilissimo metodo della spettroscopia per diffusione Brillouin, dal nome del fisico francese Louis Marcel Brillouin. Come la spettroscopia Raman, la spettroscopia per diffusione Brillouin eseguita su campioni ad alte pressioni dipende dalla trasparenza del contenitore del campione, e quindi ancora una volta la ricerca è stata resa possibile dalla cella a incudini di diamante.

Sul campione di idrogeno solido viene focalizzata luce laser monocromatica e parte della luce viene riflessa in tutte le direzioni dalle onde sonore del campione. In qualsiasi direzione la frequenza del fascio di luce riflesso viene modificata dal moto dell'onda sonora dalla quale viene riflesso: la distanza tra le creste successive dell'onda luminosa riflessa aumenta se l'onda sonora si sta allontanando dal fascio incidente, mentre diminuisce se l'onda sonora si sta avvicinando al fascio incidente. Il fenomeno è il ben noto effetto Doppler. La velocità di un'onda sonora in un solido dipende dalla densità del mezzo. Misurando perciò lo spostamento di frequenza della luce riflessa dal campione di idrogeno pressurizzato, si può calcolare la densità dell'idrogeno solido alla pressione applicata. È poi immediato ricavare il volume dell'idrogeno dalla sua densità.

I nuovi dati ottenuti dai due gruppi di ricercatori consentono di estrapolare la pressione necessaria per la transizione alla fase metallica dell'idrogeno con maggiore sicurezza che in precedenza. Le stime migliori fanno pensare che la transizione debba avvenire a pressioni comprese tra due e sei milioni di atmosfere, a seconda dei fattori di correzione che intervengono nei calcoli. Non si sa ancora se la transizione può essere osservata nella cella a incudini di diamante.

Tra le applicazioni più importanti degli studi eseguiti con la cella a incudini di diamante vi sono quelle in geofisica e in planetologia. Per lo scienziato sperimentale, che deve accontentarsi di campioni quasi microscopici di materiale sottoposto ad alta pressione, è quasi ironico che sotto di noi esista un gigantesco laboratorio delle alte pressioni. All'interno della Terra e degli altri pianeti sia la pressione sia la temperatura aumentano con la profondità e al centro della Terra la pressione è di quattro milioni di atmosfere, mentre la temperatura è di 4000 gradi centigradi. Per capire l'origine e la dinamica di qualsiasi pianeta è necessario conoscere nei dettagli le variazioni di densità, di fase e di composizione chimica del pianeta con la profondità, e tutti questi fattori sono influenzati da variazioni di pressione e di temperatura. Soprattutto sulla base di dati sismici, si sa che l'interno della Terra è diviso in un nucleo, un mantello e una crosta. Sarebbe tuttavia interessante poter simulare le condizioni di pressione e di temperatura a varie profondità in un pianeta per poter osservare direttamente i loro effetti sui materiali dai quali si pensa sia composto il pianeta.

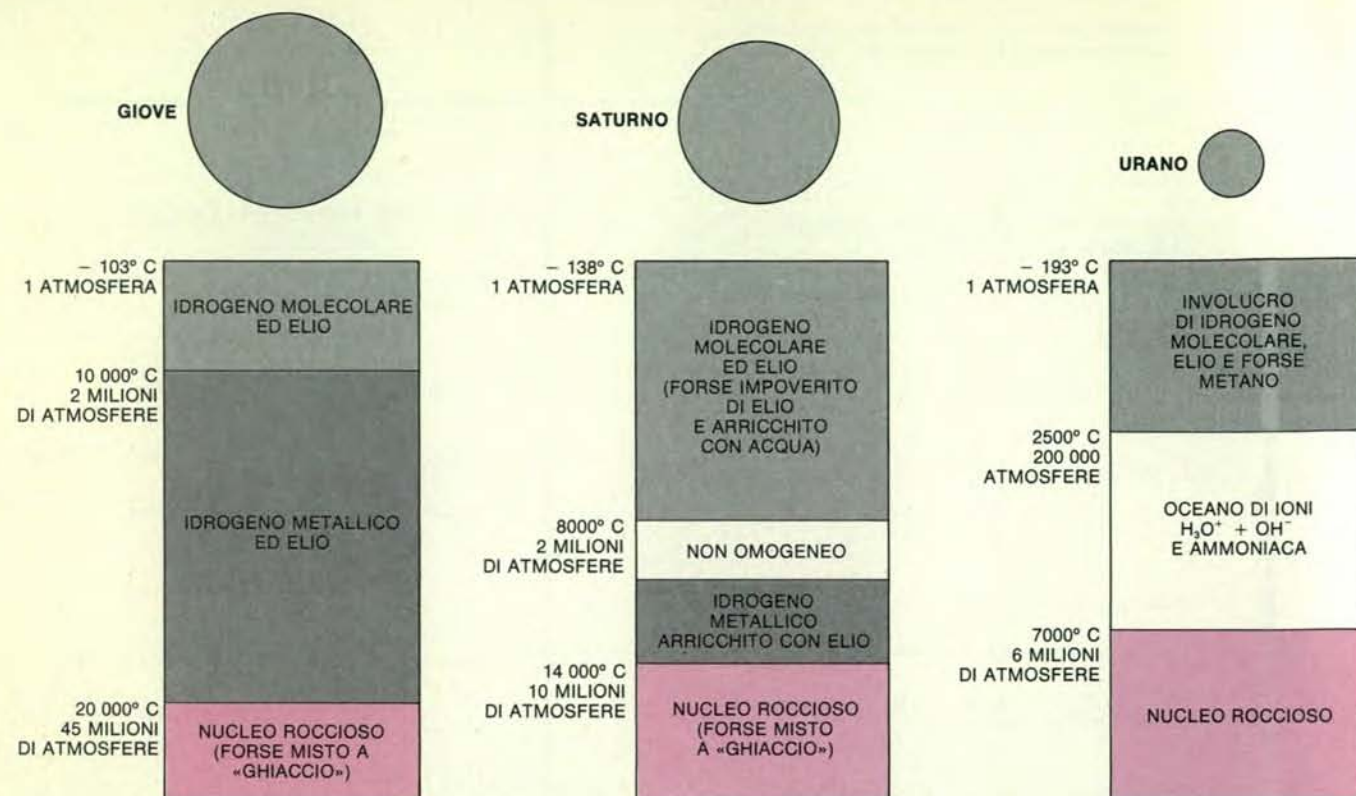
Prima della realizzazione della cella a incudini di diamante studi diretti di questo genere erano possibili solo per condizioni di pressione e di temperatura corrispondenti a una profondità di circa 400 chilometri. Si incontravano difficoltà non solo nella simulazione di pressioni più elevate, ma anche nel riscaldamento del materiale pressurizzato a una determinata temperatura. La trasparenza del diamante è stata sfruttata per ovviare al problema del riscaldamento. All'Università di Rochester, Li-chung Ming e William A. Bassett hanno dimostrato che, focalizzando sul campione, attraverso un diamante, il fascio di un laser a ittrio-alluminio-granato (YAG), si poteva riscaldare rapidamente il campione fino a circa 2000 gradi centigradi. Si può misurare la temperatura osservando il colore del campione riscaldato. Perciò, nelle mani del geofisico, la cella a incudini di diamante è diventata una finestra sull'interno della Terra in condizioni corrispondenti a quelle che sono rinvenibili a 2900 chilometri di profondità dalla superficie, ossia alla superficie di separazione tra mantello e nucleo.

PRESSIONE (ATMOSFERE)

10^{-32} — «Pressione» di non equilibrio dell'idrogeno gassoso nello spazio intergalattico
10^{-28}
10^{-24} — «Pressione» di non equilibrio della radiazione cosmica di fondo a microonde
10^{-20} — Pressione nello spazio interplanetario
10^{-16} — Il vuoto più spinto ottenuto in laboratorio
10^{-12} — Pressione atmosferica a 500 chilometri di altezza
— Pressione di un'onda sonora alla soglia di udibilità
10^{-8} — Pressione della luce solare intensa alla superficie della Terra
10^{-4} — Pressioni statiche misurabili in laboratorio sulla Terra
1
10^{4}
10^{8} — Pressione alla quale il diamante diventa carbonio metallico
— Pressione al centro di Giove
— Pressione alla quale l'elio diventa un solido metallico
10^{12} — Pressione di radiazione al centro del Sole
— Pressione al centro del Sole
10^{16} — Pressione al centro di una stella gigante rossa
— Pressione al centro di una stella nana bianca
10^{20}
10^{24} — Pressione al centro di una stella superdensa
10^{28} — Pressione al centro di una stella di neutroni
10^{32}

PRESSIONE (ATMOSFERE)

10^{-7} — Pressione parziale dell'idrogeno nell'atmosfera al livello del mare
10^{-6} — Il vuoto più spinto ottenibile con una pompa meccanica
— Pressione di radiazione sulla superficie del Sole
10^{-5} — Pressione della zampa di un gerride sulla superficie dell'acqua
10^{-4} — Pressione osmotica del saccarosio alla concentrazione di 1 milligrammo per litro
— Pressione di un'onda sonora alla soglia del dolore (120 decibel)
10^{-3} — Pressione parziale dell'anidride carbonica nell'atmosfera al livello del mare
— Tensione di vapore dell'acqua al punto triplo dell'acqua
10^{-2} — Sovrapressione nella bocca prima dell'emissione di una consonante *p*
— Pressione atmosferica all'altezza di 30 chilometri
— Pressione entro una lampadina
10^{-1} — Sovrapressione media del sangue arterioso nell'uomo
— Pressione atmosferica sulla cima dell'Everest
1 — Pressione atmosferica al livello del mare
— Pressione di un pattinatore sul ghiaccio
10 — Pressione massima all'interno di un cilindro di un motore ad alta compressione
— Pressione dell'aria in un pneumatico di bicicletta ad alta pressione
— Pressione del vapore nella caldaia di una centrale di potenza
10^{2} — Pressione dell'aria in una bombola da sommozzatore
— Pressione dello stilo su un disco fonografico
10^{3} — Pressione massima di un pugno sul cemento in un colpo di karatè
— Pressione alle massime profondità oceaniche
— Pressione di solidificazione del mercurio a temperatura ambiente
10^{4} — Pressione alla quale l'acqua congela a 100°C
— Pressione alla quale la grafite diventa diamante
— Pressione massima ottenibile in laboratorio prima dell'impiego della cella a incudini di diamante
10^{5} — Pressione alla quale il ferro diventa non magnetico
— Transizione da spinello a perovskite, a 670 chilometri di profondità dalla superficie della Terra
— Pressione di radiazione di un fascio focalizzato di luce laser a impulsi
10^{6} — Massima pressione ottenuta con una cella a incudini di diamante
— Pressione alla quale l'idrogeno diventa un solido metallico
— Pressione al centro della Terra
10^{7} — Pressione al centro di Saturno

La gamma delle pressioni nell'universo è rappresentata graficamente per più di 60 ordini di grandezza su due scale logaritmiche. La scala a destra è una versione ingrandita della parte centrale della scala a sinistra. La cella a incudini di diamante ha permesso di ampliare di oltre un ordine di grandezza il precedente limite delle alte pressioni statiche studiabili. Le transizioni alle pressioni più alte sono previste dalla teoria.

La composizione interna dei tre pianeti giganti viene ricavata da calcoli teorici basati, quando possibile, sull'estrapolazione di dati sperimentali. Le ricerche eseguite sullo iodio solido ad alta pressione fanno pensare che anche l'idrogeno molecolare solido subisca analoghe transizioni di fase e diventi un metallo conduttore a una pressione compresa tra due e sei milioni di atmosfere. Si ipotizza perciò che nell'interno dei pianeti, alle profondità corrispondenti a tali pressioni, l'idrogeno si trovi in una forma metallica. L'ulteriore differenziazione delle strutture planetarie si basa su studi spettroscopici e su estrapolazioni teoriche. I «ghiacci» presenti nel nucleo di Giove e in quello di Saturno sono ammoniaca solida, metano solido e ghiaccio comune. I modelli sono stati realizzati da David J. Stevenson del California Institute of Technology.

Una scoperta fondamentale è che i silicati formati da biossido di silicio (SiO_2), ossido di magnesio (MgO), ossido di alluminio (Al_2O_3) e ossido di ferro (FeO) subiscono una serie di cambiamenti di fase indotti dalla pressione. Ogni cambiamento di fase provoca una variazione di densità e le variazioni di densità si possono correlare con le variazioni osservate della velocità delle onde sismiche con la profondità. È stato possibile correlare con i dati sismici due importanti transizioni di fase. Un improvviso aumento di velocità delle onde sismiche a una profondità di circa 400 chilometri si ritiene sia dovuto a un cambiamento di fase dell'olivina, un silicato di magnesio e ferro [$(Mg,Fe)_2SiO_4$], in una struttura più densa chiamata spinello. Allo stesso modo, un aumento di velocità delle onde sismiche a una profondità di 670 chilometri è associato a una trasformazione da spinello alla struttura chiamata perovskite. La fase della perovskite sembra stabile e si crede che sia predominante in tutto il mantello inferiore. Esistono altre anomalie nella velocità delle onde sismiche dovute probabilmente anch'esse a cambiamenti di fase. Tali cambiamenti potrebbero essere indotti anche da silicati meno abbondanti tra i quali l'ossido di calcio (CaO) o l'ossido di alluminio.

Le condizioni prevalenti in tutto il corpo dei pianeti giganti non sono ancora direttamente simulabili con la cella a incudini di diamante. Per esempio, si stima che al centro di Giove la pressione sia di circa 45 milioni di atmosfere, mentre al centro di Saturno è probabilmente di 10 milioni di atmosfere: sono pressioni di un ordine di grandezza superiori alla massima registrata finora con la cella a incudini di diamante. Ciononostante, gli esperimenti ad alta pressione con materiali quale l'idrogeno consentono di costruire modelli teorici più attendibili della composizione e della dinamica dell'interno dei pianeti.

Si ritiene che l'idrogeno sia il materiale dominante sia su Giove sia su Saturno, e si ipotizza che esso sia presente in forma metallica a profondità superiori a quella corrispondente alla transizione alla fase metallica. Nel nucleo dei due pianeti possono esserci ossidi minerali di magnesio, silicio e ferro e perfino tre tipi di «ghiaccio»: ammoniaca solida, metano solido e ghiaccio d'acqua. I ghiacci sono probabilmente solo costituenti minoritari di Giove e di Saturno, mentre si ritiene siano importanti nel nucleo sia di Urano sia di Nettuno.

A questo punto nasce spontanea una domanda: qual è la massima pressione ottenibile nella cella a incudini di diamante? Un primo limite dovrebbe essere imposto da una transizione di fase nello stesso diamante, forse a una fase metallica. Un calcolo recente di Mingtang Yin e Marvin L. Cohen dell'Università della California a Berkeley mostra che la struttura del diamante non si riduce ad alcuna delle sette ben note strutture metalliche cristalline a pressioni inferiori a 23 milioni di atmosfere. Inoltre, si è generalmente dell'avviso che il diamante sia stabile rispetto a qualsiasi altra transizione metallica fino a pressioni di almeno 10 milioni di atmosfere. Sfortunatamente è probabile che il diamante subisca deformazioni plastiche a pressioni inferiori alla transizione alla fase metallica, e quindi la massima pressione ottenibile con la cella a incudini di diamante sarà probabilmente determinata dalla resistenza allo snervamento.

Dalla teoria della plasticità è noto che la resistenza allo snervamento del diamante è massima in quelle regioni nelle quali è perfetta la sua struttura cristallina. Per esempio, si calcola che in un diamante perfetto si debbano raggiungere pressioni fino a cinque milioni di atmosfere prima che il diamante inizi a snervarsi. Minore è la regione del diamante che deve sopportare le più alte pressioni, maggiore è la probabilità che la regione abbia una struttura cristallina perfetta. Pare quindi che il modo di ottenere maggiori pressioni sia di eseguire esperimenti su una scala progressivamente miniaturizzata.

La visione nell'uomo e nella macchina

In che modo vede l'uomo e in che modo potrebbe vedere un calcolatore? Gli studi di neurobiologia suggeriscono i programmi per il calcolatore il quale a sua volta suggerisce che cosa si dovrebbe cercare nel cervello

di Tomaso Poggio

L'uso di calcolatori sempre più potenti e raffinati per risolvere problemi un tempo riservati esclusivamente all'uomo rende sempre più attuale il confronto tra queste macchine e il cervello umano. In effetti una futura generazione di robot e di calcolatori potrebbe essere dotata di capacità sensoriali, motorie e «intellettuali» simili alle nostre. Come costruire macchine simili? Da una parte ci possono fare da guida le nostre sempre crescenti conoscenze sul cervello. D'altra parte, e al tempo stesso, i progressi scientifici nel campo dell'intelligenza artificiale ci insegneranno a capire il cervello.

A livello delle strutture e dei meccanismi, ossia a livello di *hardware*, cervello e calcolatori sono profondamente diversi. I neuroni, le cellule nervose del cervello, sono strutture piccole e delicate, assai diverse dai conduttori e semiconduttori su cui sono basati i calcolatori di oggi. Anche a livello dell'organizzazione dello hardware vi sono enormi differenze: i collegamenti tra i neuroni sono numerosissimi (ogni neurone può ricevere parecchie migliaia di segnali) e sono distribuiti tridimensionalmente; i conduttori che collegano i componenti di un odierno calcolatore sono invece relativamente pochi e distribuiti bidimensionalmente. Un'altra fondamentale differenza è costituita dalla trasmissione dei segnali. Gli impulsi elettrici binari del calcolatore *(on-off)* sono in parte paragonabili ai segnali «tutto o nulla» trasmessi lungo le cellule nervose. Il cervello però per trasmettere informazioni utilizza anche segnali graduati, sostanze chimiche e il trasporto di ioni. Se si considera l'organizzazione temporale le differenze sono immense. Mentre i calcolatori elaborano l'informazione passo per passo, serialmente, ad altissima velocità, regolati da un sincronizzatore *(clock)* centrale, il cervello analizza l'informazione più lentamente su milioni di canali contemporaneamente, senza bisogno di un sincronizzatore generale.

In che senso allora possiamo paragonare un calcolatore al cervello? C'è un livello al quale paragonare due diversi meccanismi. Se consideriamo per esempio i calcolatori e il cervello come dei sistemi che elaborano informazione, possiamo paragonare e descrivere i compiti svolti da entrambi. A questo punto abbiamo un linguaggio comune: il linguaggio dell'elaborazione dell'informazione. È importante notare però che il linguaggio con cui si descrive lo svolgimento di determinati compiti è indipendente dalla struttura «fisica», dallo hardware, che svolge tali compiti. Questa separabilità è uno degli aspetti fondamentali della scienza dell'intelligenza artificiale, i cui obiettivi sono da un lato quello di rendere più utili i calcolatori, dotandoli di capacità «intelligenti», e dall'altro quello di capire quali siano i principi che rendono possibile l'intelligenza. Lo studio di come vediamo rappresenta uno dei campi dell'intelligenza artificiale in cui si hanno le descrizioni più precise di problemi di elaborazione dell'informazione. Una delle ragioni è naturalmente che la visione rappresenta la modalità sensoriale dominante nell'uomo. Se vogliamo creare robot capaci di eseguire compiti manipolativi complessi in un ambiente in continuo mutamento, dobbiamo senz'altro dotarli di adeguate capacità visive. Le più precise descrizioni di compiti di elaborazione dell'informazione hanno infatti come oggetto compiti visivi. Si tratta di un problema apparentemente semplice, in quanto vediamo senza alcuno sforzo, quasi senza accorgercene, facilmente e rapidamente. Di conseguenza tendiamo a semplificare, sottovalutare e minimizzare i problemi connessi con la visione. Se per vedere dovessimo compiere uno sforzo cosciente, come quando per esempio sommiamo due numeri, non cadremmo in questo errore. In effetti la visione rappresenta un problema estremamente complesso, dal punto di vista computazionale, con livelli di elaborazione di informazione di dimensioni mastodontiche. Sta di fatto che, per creare robot capaci di svolgere compiti manipolativi di una certa complessità in un ambiente in continuo mutamento, dobbiamo prima sapere come dotarli di capacità visive.

A livello biologico ci serve capire come i neuroni e le sinapsi «elaborino» l'informazione visiva, ma non si tratta in realtà soltanto di un problema anatomico e fisiologico. Dal punto di vista dell'elaborazione dell'informazione sia da parte del cervello sia da parte di un calcolatore si tratta di un problema a molti livelli: il livello del calcolo (quali compiti computazionali

La visione stereoscopica da parte di un calcolatore esemplifica lo studio della visione come problema di elaborazione di informazioni. Le immagini in alto nella pagina a fronte sono fotografie aeree fornite da Robert J. Woodham dell'Università della British Columbia e mostrano parte del campus universitario. Riproducono in due modi i dati visivi su cui si basa la visione biologica: in primo luogo sono state scattate da angolazioni diverse, cosicché gli oggetti in un'immagine hanno una posizione leggermente diversa rispetto a quella che hanno nell'altra, così come i due occhi di un uomo vedono il mondo da angolazioni diverse. In secondo luogo sono state prodotte da un mosaico di sensori microelettronici, ciascuno dei quali misura l'intensità di luce lungo una particolare linea di vista. I fotorecettori dell'occhio svolgono un compito molto simile. La mappa in basso è stata generata da un calcolatore programmato per seguire un algoritmo formulato da David Marr e dall'autore all'Artificial Intelligence Laboratory del Massachusetts Institute of Technology e ulteriormente sviluppato nello stesso laboratorio da W. Eric L. Grimson. Il calcolatore ha filtrato le immagini per evidenziare variazioni spaziali di intensità, poi ha operato la visione stereoscopica: ha messo in corrispondenza caratteristiche dell'una e dell'altra immagine, determinato la disparità fra le loro posizioni e calcolato la loro profondità relativa nel mondo tridimensionale. Altezze crescenti nella mappa sono codificate con le tonalità di colore dal blu verso il rosso.

225	221	216	219	219	214	207	218	219	220	207	155	136	135	130	131	125
213	206	213	223	208	217	223	221	223	216	195	156	141	130	128	138	123
206	217	210	216	224	223	228	230	234	216	207	157	136	132	137	130	128
211	213	221	223	220	222	237	216	219	220	176	149	137	132	125	136	121
216	210	231	227	224	228	231	210	195	227	181	141	131	133	131	124	122
223	229	218	230	228	214	213	209	198	224	161	140	133	127	133	122	133
220	219	224	220	219	215	215	206	206	221	159	143	133	131	129	127	127
221	215	211	214	220	218	221	212	218	204	148	141	131	130	128	129	118
214	211	211	218	214	220	226	216	223	209	143	141	141	124	121	132	125
211	208	223	213	216	226	231	230	241	199	153	141	136	125	131	125	136
200	224	219	215	217	224	232	241	240	211	150	139	128	132	129	124	132
204	206	208	205	233	241	241	252	242	192	151	141	133	130	127	129	129
200	205	201	216	232	248	255	246	231	210	149	141	132	126	134	128	139
191	194	209	238	245	255	249	235	238	197	146	139	130	132	129	132	123
189	199	200	227	239	237	235	236	247	192	145	142	124	133	125	138	128
198	196	209	211	210	215	236	240	232	177	142	137	135	124	129	132	128
198	203	205	208	211	224	226	240	210	160	139	132	129	130	122	124	131
216	209	214	220	210	231	245	219	169	143	148	129	128	136	124	128	123
211	210	217	218	214	227	244	221	162	140	139	129	133	131	122	126	128
215	210	216	216	209	220	248	200	156	139	131	129	139	128	123	130	128
219	220	211	208	205	209	240	217	154	141	127	130	124	142	134	128	129
229	224	212	214	220	229	234	208	151	145	128	128	142	122	126	132	124
252	224	222	224	233	244	228	213	143	141	135	128	131	129	128	124	131
255	235	230	249	253	240	228	193	147	139	132	128	136	125	125	128	119
250	245	238	245	246	235	235	190	139	136	134	135	126	130	126	137	132
240	238	233	232	235	255	246	168	156	141	129	127	136	134	135	130	126
241	242	225	219	225	255	255	183	139	141	126	139	128	137	128	128	130
234	218	221	217	211	252	242	166	144	139	132	130	128	129	127	121	132
231	221	219	214	218	225	238	171	145	141	124	134	131	134	131	126	131
228	212	214	214	213	208	209	159	134	136	139	134	126	127	127	124	122
219	213	215	215	205	215	222	161	135	141	128	129	131	128	125	128	127

L'inizio della visione per un animale o un calcolatore è una matrice di livelli di grigio: una rappresentazione punto per punto dell'intensità di luce, prodotta da una griglia di rivelatori nell'occhio o in una fotocamera digitale. L'immagine in alto è una matrice di questo tipo, prodotta da una fotocamera digitale come insieme di valori di intensità in una griglia di 576 per 454 pixel. I valori di intensità per la parte di immagine nel rettangolo sono dati in forma digitale in basso.

deve svolgere il sistema visivo); il livello dell'algoritmo (in quale successione di stadi elementari si completa un compito); il livello dello hardware (come potrebbero eseguire l'algoritmo i neuroni o i circuiti elettronici). Dobbiamo quindi basarci su dati di natura psicofisica (come vede l'uomo) e su dati di natura neurofisiologica (cosa fanno i neuroni) per poter affrontare il problema della visione. L'aspetto più critico riguarda gli algoritmi, perché per essere realizzabili debbono soddisfare sia vincoli imposti dai calcoli sia vincoli imposti dallo hardware. A questo punto cercherò di delineare la ricerca in cui sono direttamente impegnato. Si tratta dell'analisi di una serie di algoritmi utili a estrarre informazione, in particolare a evidenziare i contorni di immagini e quindi a calcolare, sulla base di questi contorni, la profondità degli oggetti nel mondo tridimensionale. Mi concentrerò su un aspetto particolare di tale problema, cioè sulla visione stereoscopica o binoculare. Essa ha assunto un ruolo centrale nelle ricerche sulla visione condotte dai miei collaboratori e da me all'Artificial Intelligence Laboratory del Massachusetts Institute of Technology; in particolare si presta a una indagine piuttosto dettagliata dei primissimi stadi dell'elaborazione dell'informazione visiva. Un fatto importante che rende attraente lo studio della visione stereoscopica è l'abbondanza di dati psicofisici disponibili, che definiscono e delimitano il problema.

Come ho già accennato all'inizio a proposito della visione in generale, anche la visione stereoscopica è un processo ingannevolmente semplice, che il cervello svolge senza alcuno sforzo. Si vedrà però come lo sviluppo di un sistema dotato di «visione stereoscopica» si sia rivelato estremamente difficile. L'informazione disponibile all'inizio del processo visivo è una matrice bidimensionale costituita dalle misure della quantità di luce riflessa nell'occhio o nella fotocamera da ciascun punto della superficie di oggetti tridimensionali. Nell'occhio umano oltre cento milioni di fotorecettori (coni e bastoncelli) svolgono questo compito. Nella fotocamera che abbiamo sviluppato e usiamo all'Artificial Intelligence Laboratory i processi sono diversi, ma il risultato molto simile: le misurazioni vengono fatte da sensori elettronici, che producono una matrice di 1000 per 1000 valori di intensità luminosa, ciascuno dei quali è un pixel (un *picture element*), ossia un elemento dell'immagine. È impossibile pensare di poter passare direttamente dall'immagine grezza (la matrice dei numeri prodotta dall'occhio o dalla fotocamera) alla visione (il sapere che cosa c'è nell'immagine e dove). In effetti la visione richiede vari processi («moduli») che agiscono in parallelo sull'immagine grezza e producono rappresentazioni intermedie dell'immagine stessa su cui possono operare altri processi. Numerosi sono i moduli visivi coinvolti nella ricostruzione del mondo tridimensionale. Eccone un breve elenco: moduli che deducono la forma dal chiaro-

Gli stereogrammi a punti casuali inventati da Bela Julesz agli AT&T Bell Laboratories sono figure visive che non contengono altri indizi per la visione stereoscopica al di fuori delle disparità binoculari. Gli stereogrammi in sé sono una struttura casuale di punti bianchi e neri *(in alto)*. In uno di essi un quadrato è spostato verso sinistra; nell'altro è spostato verso destra *(in basso)*. La lacuna risultante in ciascuna immagine è riempita con altri punti casuali *(zone in grigio)*.

scuro, dalla tessitura visiva, dal movimento, dai contorni, dalle occlusioni e dalla visione stereoscopica. Alcuni operano direttamente sull'immagine grezza, mentre altri agiscono con maggior efficacia su una rappresentazione intermedia.

La visione stereoscopica è dovuta al fatto che i nostri occhi vedono il mondo da angoli leggermente diversi; in altri termini gli occhi sono leggermente convergenti e i loro assi visivi si incontrano in un punto. Tale punto viene fissato dagli occhi e l'immagine del punto cade nel cosiddetto centro della visione di ciascuna retina. Qualunque altro punto nel campo visivo sarà quindi proiettato su ciascuna retina a una certa distanza dal centro della visione. Generalmente questa distanza non è la stessa per ambedue gli occhi, ma la differenza di distanza varia con la profondità del punto nel campo visivo rispetto al punto fissato. La visione stereoscopica è il processo che decodifica la tridimensionalità dalle disparità binoculari. A prima vista si potrebbe pensare di risolvere il problema come se si trattasse di un semplice problema di trigonometria. Il tentativo però fallirebbe; non si tratta di un processo semplice come appare, ma assai complesso che consiste di almeno quattro stadi: in un'immagine si sceglie la proiezione di un punto nello spazio; si identifica la posizione nell'altra immagine che corrisponde allo stesso punto; si misurano le posizioni nelle due immagini; dalla differenza tra le due misurazioni si calcola la distanza del punto.

Gli ultimi due stadi sono effettivamente un compito di trigonometria (per lo meno nei casi considerati in questo articolo), ma i primi due sono diversi. Essi richiedono l'identificazione della proiezione dello stesso punto del mondo fisico in ciascun occhio. Si può immaginare che un gruppo di fotorecettori contigui in un occhio osservi lungo una linea di vista una porzione della superficie dello stesso oggetto. Si debbono poi identificare i fotorecettori che guardano la stessa porzione di superficie dall'altro occhio. Data la disparità binoculare, non si troveranno nella stessa posizione rispetto al centro della visione.

Qui sta ovviamente la difficoltà. Noi siamo in grado di distinguere le superfici nel mondo visivo, in quanto esse fanno parte di forme diverse in specifiche relazioni spaziali l'una con l'altra. Si deve però ricordare che la visione comincia con una matrice di intensità luminose misurate punto per punto. È possibile che il cervello metta in corrispondenza configurazioni di intensità luminose da un occhio all'altro? Probabilmente no. Gli esperimenti con i calcolatori ci mostrano i limiti dell'efficacia di questa corrispondenza, e i dati fisiologici e psicofisici sono contro questa possibilità per il sistema visivo umano. Una data porzione di superficie non riflette necessariamente la stessa intensità di luce in ambedue gli occhi; cosa ancora più importante, porzioni di superficie molto distanti del mondo visivo potrebbero avere la stessa intensità, e mettere in corrispondenza tali porzioni sarebbe scorretto.

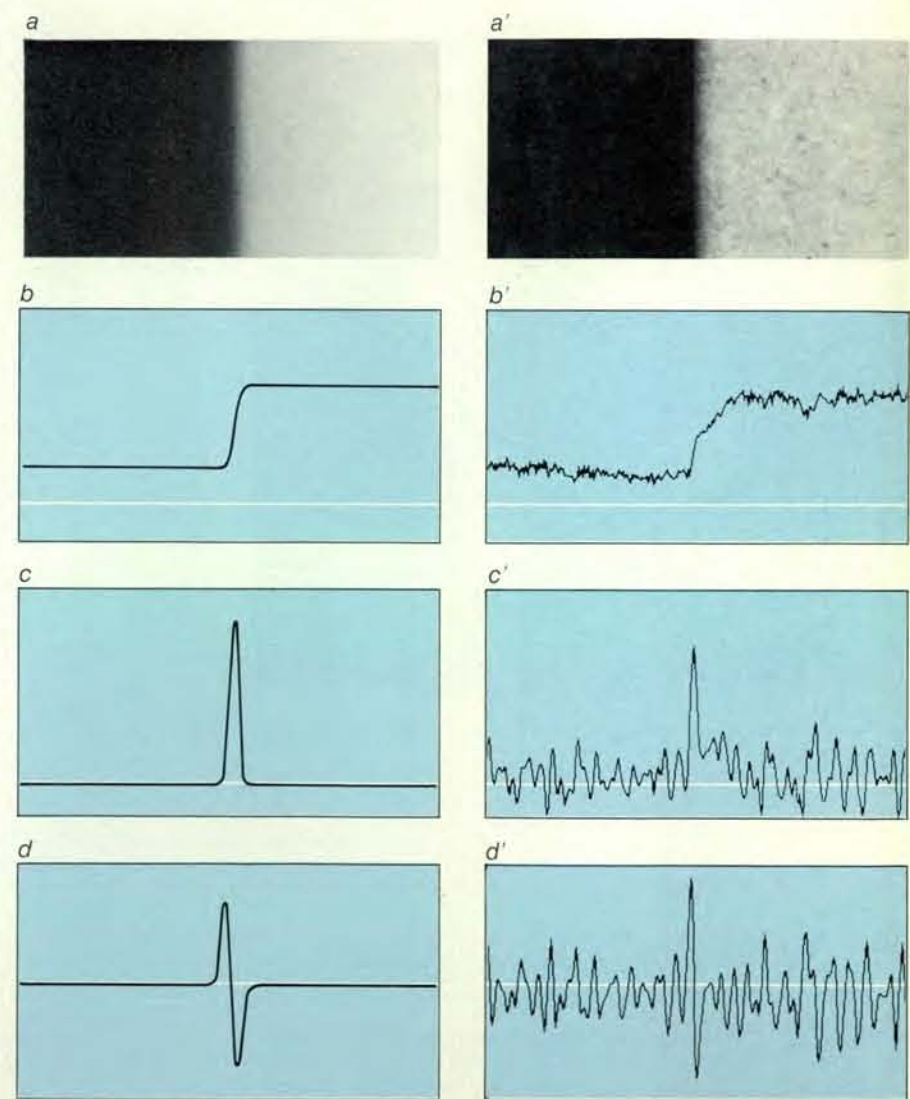
Quando gli stereogrammi a punti casuali vengono visti attraverso uno stereoscopio, in modo che ciascun occhio veda solo uno dei due stereogrammi e il cervello possa fondere le due immagini, si ha una vivida percezione di profondità. Questo fenomeno dell'immagine «fluttuante» dimostra che la visione stereoscopica non presuppone il riconoscimento di oggetti nel mondo visivo.

Una scoperta di Bela Julesz (AT&T Bell Laboratories) mostra tutta la complessità del problema. Julesz ha prodotto quelli che ha chiamato *random dot stereograms* (stereogrammi a punti casuali). Si tratta di figure costituite da punti bianchi e neri distribuiti a caso e che contengono disparità binoculari. Ci sono due copie di ogni figura. In una copia una porzione della configurazione, per esempio un quadrato, viene leggermente spostata da una parte, mentre nell'altra un'egual porzione viene spostata in direzione opposta. La porzione rimasta vuota in ciascuna configurazione viene riempita con altri punti a caso. Viste una alla volta le due copie di una figura sembrano costituite da punti uniformemente casuali. Viste invece attraverso uno stereoscopio, in modo che ciascun occhio veda una delle due copie della figura e il cervello fonda le due immagini, il risultato è sorprendente: si ha la nettissima impressione che il quadrato si stacchi dallo sfondo della figura. Evidentemente la visione stereoscopica non richiede il riconoscimento di oggetti visti precedentemente o di forme.

Questa scoperta di Julesz ci permette di formulare i compiti computazionali coinvolti nella visione stereoscopica: l'estrazione di disparità binoculari da una coppia di immagini non richiede ovvi indizi monoculari; per mettere in rapporto gli elementi corrispondenti delle due immagini non c'è bisogno di riconoscere gli oggetti o le loro parti (problema della corrispondenza). Negli stereogrammi a punti casuali, i punti neri di ciascuna immagine sono del tutto identici: hanno le stesse dimensioni, forma e luminosità. In linea di principio ciascun punto di una immagine potrebbe essere messo in corrispondenza con un gran numero di punti dell'altra immagine. Il cervello tuttavia risolve questo dilemma scegliendo coerentemente solo l'insieme corretto di corrispondenze.

Esso deve usare qualcosa di più dei punti stessi. In particolare, il fatto che il cervello risolva il problema della corrispondenza indica che è in grado di sfruttare con efficacia un insieme di ipotesi implicite sul mondo visivo, che definiscono il problema della corrispondenza rendendolo risolvibile. Nel 1976 David Marr e io, lavorando al Massachusetts Institute of Technology, abbiamo trovato che alcune semplici proprietà delle superfici fisiche potevano delimitare il problema a sufficienza per gli algoritmi (cioè le procedure che un calcolatore deve seguire) di visione stereoscopica che stavamo studiando. In primo luogo, un punto dato su una superficie fisica ha solo una localizzazione tridimensionale in ciascun istante; secondariamente, gli oggetti fisici sono compatti e di solito opachi, cosicché la variazione di profondità su una superficie in genere è continua, con discontinuità solo in corrispondenza dei bordi. Il primo di questi vincoli, l'unicità di localizzazione, significa che ciascun elemento di ciascuna immagine (per esempio ciascun puntino in uno stereogramma a punti casuali) ha un'unica disparità e può essere messo in corrispondenza con non più di un elemento nell'altra immagine. Il secondo vincolo, continuità e opacità, significa che la disparità varia con continuità, fatta eccezione per i confini dell'oggetto.

Insieme i due vincoli forniscono regole di corrispondenza ragionevoli e potenti. Ne descriverò alcune semplici più avanti. Prima però è necessario specificare gli elementi da mettere in corrispondenza. Dopo tutto, il mondo visivo non è uno stereogramma a punti casuali, costituito solo da punti bianchi e neri. Abbiamo già visto che i valori di intensità dell'immagine sono troppo ambigui. Tuttavia le informazioni di cui il cervello ha bisogno sono cifrate nella matrice di intensità fornita dai fotorecettori. Se si chiama in causa un'ulteriore proprietà delle superfici fisiche, il problema si semplifica. Nei punti dove vi sono variazioni fisiche in una superficie, l'immagine della superficie di solito presenta nette variazioni di intensità. Queste variazioni, causate dalla tessitura della superficie e da variazioni della sua distanza, forniscono indicazioni molto più sicure ed efficaci per la corrispondenza che non i valori di intensità.

È possibile identificare le variazioni di intensità in una immagine confrontando valori di intensità adiacenti: se la differenza tra di essi è grande, l'intensità varia rapidamente. In termini matematici l'operazione corrisponde a stimare la derivata prima. La derivata prima è semplicemente il tasso di variazione di una funzione matematica. Nel nostro caso è semplicemente la rapidità con cui l'intensità varia lungo un percorso attraverso la matrice di livelli di grigio. La posizione di un valore estremo, un massimo o un minimo, nella derivata prima identifica la posizione di un contorno nell'immagine *(si veda l'illustrazione in questa pagina)*. A sua volta questo contorno corrisponde spesso a un contorno su una superficie fisica. Anche la derivata seconda ci aiuta. È semplicemente la rapidità di cambiamento della rapidità di cambiamento e si ottiene considerando le differenze tra i va-

Le derivate spaziali di un'immagine servono per evidenziarne le variazioni di intensità nello spazio. La parte a sinistra dell'illustrazione mostra un contorno tra due tonalità uguali di grigio *(a)*. L'intensità lungo un percorso che interseca il contorno è riportata sotto *(b)*. La derivata prima dell'intensità è il tasso a cui varia l'intensità *(c)*. Verso destra o verso sinistra non c'è variazione: la derivata prima pertanto si annulla. Lungo il contorno stesso, però, il tasso di variazione sale e poi scende. La derivata seconda dell'intensità è il tasso di variazione del tasso di variazione *(d)*. Ambedue le derivate rafforzano il contorno. Nella derivata prima questo è evidenziato come un picco; nella derivata seconda è rappresentato da un passaggio attraverso zero. La parte a destra dell'illustrazione mostra un contorno più tipico del mondo visivo *(a')*. Il relativo contorno d'intensità *(b')* e le sue derivate prima e seconda *(c', d')* sono affette da «rumore». Il contorno deve essere reso regolare prima di calcolarne le derivate. Questa illustrazione e quella a pagina 70 sono state preparate da H. Keith Nashihara dell'Artificial Intelligence Laboratory.

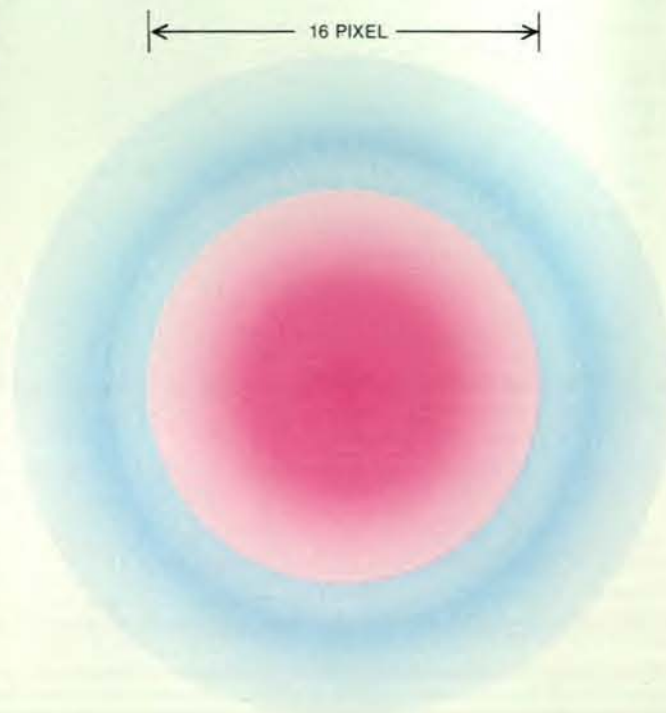
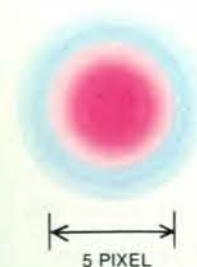
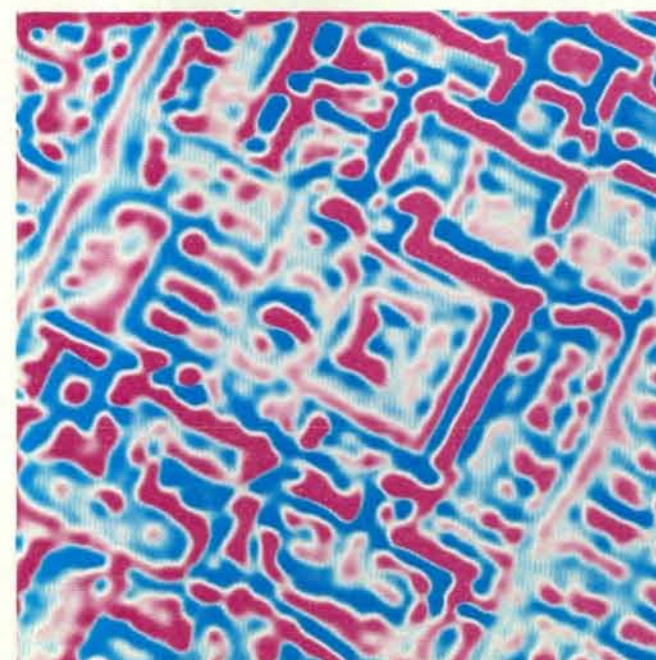
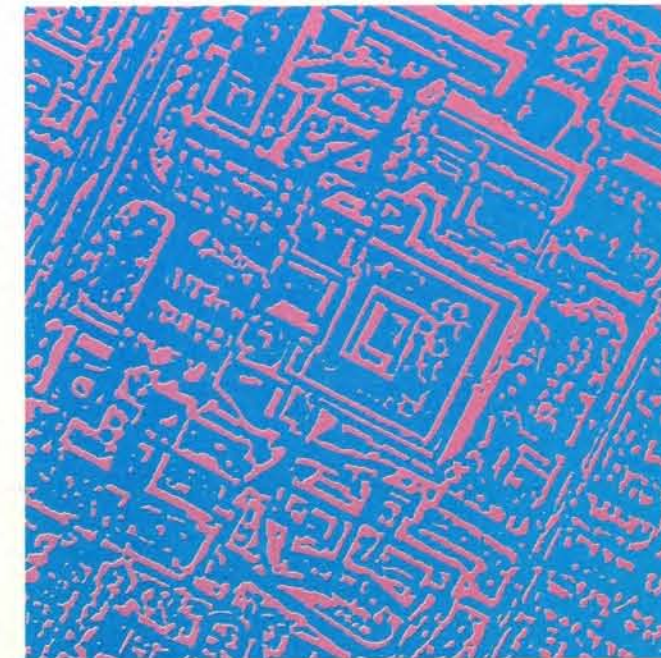
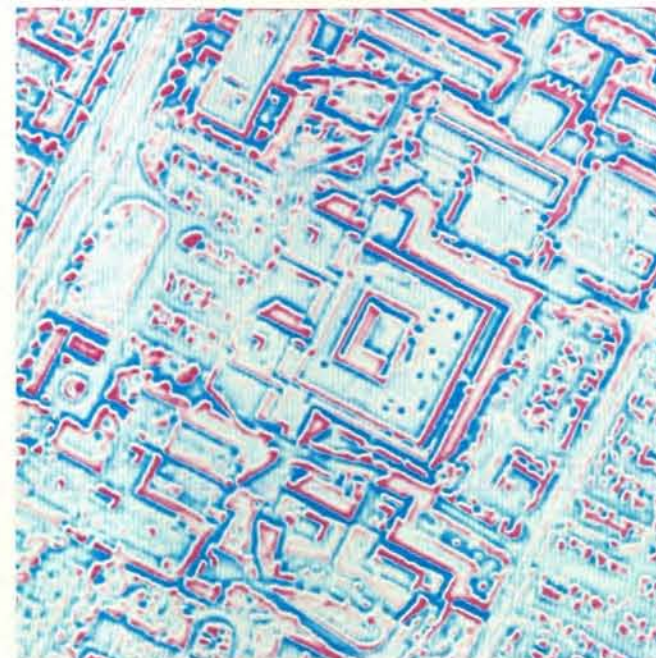
lori adiacenti della derivata prima. Nella derivata seconda un contorno di intensità nella matrice dei livelli di grigio corrisponde a un «passaggio attraverso zero», cioè a una intersezione con l'asse delle ascisse: un punto in cui la derivata seconda assume il valore zero passando da valori positivi a valori negativi, o viceversa.

Le derivate sembrano uno strumento molto promettente. Usate da sole, tuttavia, non danno buoni risultati su un'immagine reale, specialmente perché le variazioni di intensità in un'immagine reale raramente sono variazioni chiare e distinte da un valore di intensità a un altro. Spesso molti cambiamenti diversi, lenti e veloci, si intrecciano, a diverse scale spaziali. Inoltre le variazioni di intensità sono spesso mascherate dall'analogo visivo del rumore; sono mascherate, in altre parole, da fluttuazioni casuali che si infiltrano a stadi diversi quando l'immagine formata dall'ottica dell'occhio o di una fotocamera viene trasformata in una matrice di misurazioni di intensità. La ragione matematica per questa situazione è che il problema di derivare i contorni in un'immagine è un problema «mal posto». Per poter risolvere sia il problema dei contorni creati dal rumore sia il problema dei contorni su scale spaziali diverse l'immagine deve essere «regolarizzata» filtrando i valori di intensità contigui. A questo punto si può effettuare la differenziazione che corrisponde a ricavare le derivate prima e seconda.

Vi sono vari modi per effettuare questa successione di operazioni e molto lavoro è stato dedicato alla ricerca di metodi ottimali. In uno dei più semplici le due operazioni, regolarizzazione e differenziazione, sono combinate in una sola. I termini tecnici possono suonare difficili: l'immagine viene sottoposta a una convoluzione con un filtro che incorpora una particolare funzione centro-periferia, il laplaciano di una distribuzione gaussiana. Il metodo non è poi così difficile come sembra. Una gaussiana in due dimensioni è la tipica distribuzione a campana familiare agli statistici. In questo contesto specifica l'importanza che si deve assegnare alla periferia di ciascun pixel quando l'immagine viene regolarizzata. Con il crescere della distanza diminuisce l'importanza. Il laplaciano è una derivata seconda che assegna lo stesso peso a tutti i percorsi che si dipartono da un punto. Il laplaciano di una gaussiana trasforma la distribuzione a campana in una distribuzione simile a un sombrero messicano. La campana si restringe e ai lati si sviluppa una falda circolare negativa.

Ora possiamo descrivere il procedimento in termini non tecnici. Sottoporre un'immagine a una convoluzione con un filtro che incorpora il laplaciano di una gaussiana equivale a sostituire a ciascun pixel nell'immagine una media ponderata dei pixel vicini, dove i pesi sono forniti dal laplaciano di una gaussiana. Il filtro quindi viene applicato a ciascun pixel e assegna il massimo peso positivo a quel pixel e pesi positivi decrescenti ai pixel vicini *(si veda l'illustrazione in queste due pagine)*. Poi viene un anello ai cui pixel si attribuiscono pesi negativi. I punti luminosi nella figura dell'anello apportano un contributo negativo alla media. Il risultato del filtraggio complessivo è una matrice di numeri positivi e negativi: una sorta di derivata seconda dell'intensità dell'immagine alla scala del filtro. In questa matrice filtrata i passaggi attraverso zero corrispondono a punti dell'immagine originale in cui l'intensità varia più rapidamente. Si noti che una mappa binaria (cioè a due valori) che mostri semplicemente le regioni positive e negative della matrice filtrata è essenzialmente equivalente a una mappa dei passaggi attraverso zero, in quanto l'una può essere ricostruita dall'altra.

Sorprendentemente sembra che nel cervello umano sia presente la maggior parte dello hardware necessario per eseguire un simile filtraggio. Già nel 1865 Ernst Mach aveva osservato che la percezione visiva sembra rinforzare le variazioni spaziali dell'intensità luminosa. Mach speculò che ciò avvenisse mediante l'inibizione laterale, un meccanismo nervoso in cui l'eccitazione di una fibra nervosa, per esempio causata da un punto di luce intensa nel campo visivo, blocca l'eccitazione degli assoni vicini. Questa operazione intensifica il contrasto fra il punto luminoso e ciò che lo circonda, ed è analoga all'ottenere una derivata spaziale.

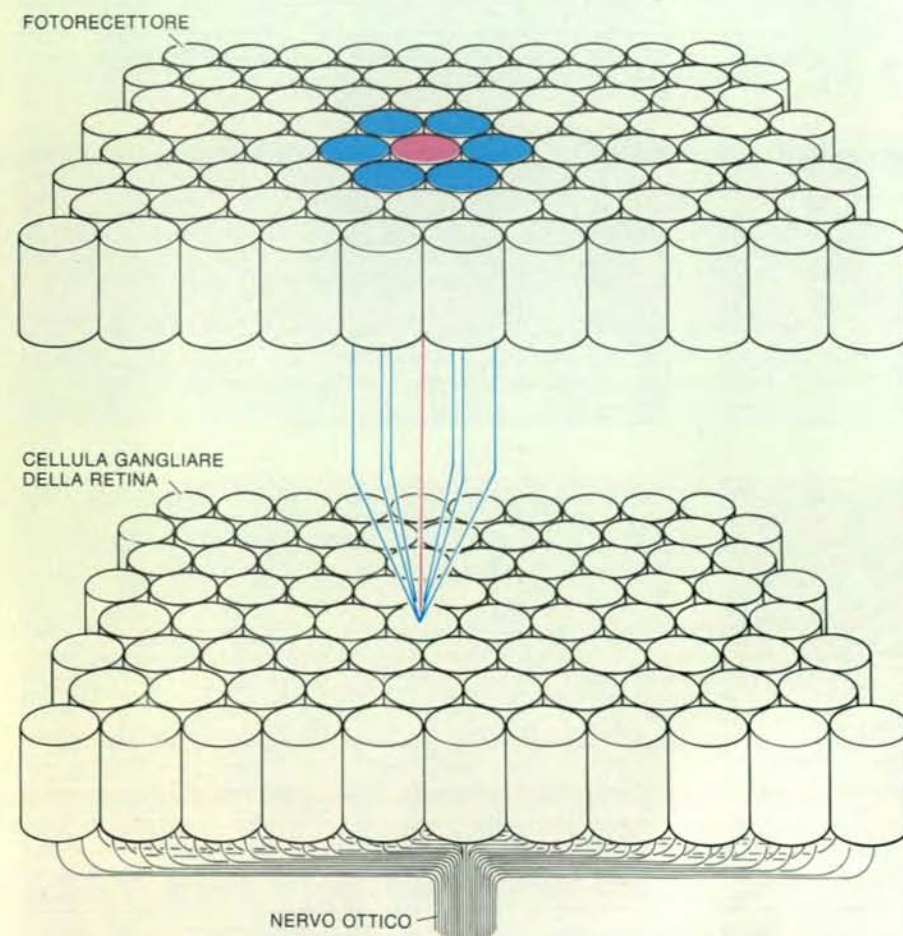
Negli anni cinquanta e sessanta si sono accumulati dati che fanno pensare che la retina compia qualcosa di simile al filtraggio centro-periferia. Il segnale di uscita di ogni retina è trasportato al resto del cervello da circa un milione di fibre nervose, ciascuna delle quali è l'assone di un neu-

Il filtraggio centro-periferia di un'immagine serve sia a regolarizzarla, sia a calcolarne la derivata seconda rispetto allo spazio. A sinistra è riportata un'immagine, a destra lo schema di due filtri (i «filtri» in realtà sono puramente computazionali). Ogni misurazione di intensità dell'immagine viene sostituita da una media ponderata delle misurazioni circostanti. Le misurazioni vicine contribuiscono alla media con pesi positivi; pertanto il centro del filtro è «eccitatorio» *(in rosso)*. L'anello in cui le misurazioni contribuiscono con pesi negativi alla «periferia» del filtro è inibitorio *(in blu)*. La terza parte dell'illustrazione mostra le mappe prodotte dai filtri. Non si tratta più di matrici di livelli di grigio: le mappe hanno sia valori positivi *(in rosso)* sia valori negativi *(in blu)*. Sono mappe della derivata seconda. Le transizioni da un colore all'altro sono passaggi attraverso zero: cioè contrassegnano le zone, nell'immagine di partenza, in cui l'intensità varia più rapidamente. Le mappe a destra nell'illustrazione mettono in evidenza i passaggi attraverso zero mostrando solo regioni positive *(in rosso)* e regioni negative *(in blu)*.

rone, chiamato cellula gangliare della retina. La cellula ottiene il suo ingresso (attraverso neuroni intermedi) da un gruppo di fotorecettori, che costituiscono un «campo recettivo». Quello che i dati sperimentali suggeriscono è che per certe cellule gangliari il campo recettivo abbia una organizzazione centro-periferia che si avvicina molto al laplaciano di una gaussiana. La luminosità al centro del campo recettivo eccita la cellula gangliare; la luminosità di un anello periferico la inibisce. In breve il campo recettivo ha un centro «ON» e una periferia «OFF», proprio come il sombrero.

Altre cellule gangliari hanno proprietà opposte: un centro «OFF» e una periferia «ON». Se gli assoni potessero trasmettere come segnali numeri negativi, queste cellule sarebbero ridondanti: esse riferiscono semplicemente la negazione di quel che riferiscono le cellule del centro «ON». I neuroni però non possono trasmettere una attività negativa: quelli che trasmettono una attività «tutto o nulla» sono o attivi o a riposo. Il cervello ha quindi bisogno di opposti neuronali. I valori positivi di un'immagine soggetta a filtraggio centro-periferia potrebbero essere rappresentati dall'attività di cellule a centro «ON», i valori negativi potrebbero essere rappresentati dall'attività di cellule a centro «OFF». A questo proposito val la pena ricordare che nella retina del gatto le cellule gangliari a centro «ON» e a centro «OFF» sono distribuite in due strati separati. Le mappe generate dal nostro calcolatore ci danno un quadro abbastanza preciso dell'attività nervosa. Nelle mappe della pagina 75 il colore rosso corrisponde all'attività dello strato ON e il blu all'attività dello strato OFF. I passaggi attraverso zero, cioè le transizioni da un colore all'altro, rappresenterebbero le localizzazioni in cui l'attività nervosa cambia da uno strato all'altro. Ecco quindi una congettura che collega una teoria computazionale della visione con lo hardware del cervello che entra in gioco nella visione biologica.

Per un calcolatore il filtraggio centro-periferia di una immagine ha un elevato costo computazionale, perché comporta un gran numero di moltiplicazioni: circa un miliardo per un'immagine di mille per mille pixel. All'Artificial Intelligence Laboratory H. Keith Nishihara e Noble G. Larson, Jr., hanno progettato un dispositivo specializzato: un «convolutore» che esegue l'operazione in circa un secondo. La velocità è notevole, ma il processo è laborioso e lento, se confrontato con quello delle cellule gangliari della retina.

Un altro importante problema è quello della scala spaziale. In un'immagine vi sono variazioni di intensità sia fini sia grossolane. Tutte debbono essere identificate e rappresentate, ma come? La soluzione naturale - e quella suggerita dalla fisiologia e dalla psicofisica - sta nell'uso di filtri centro-periferia di dimensioni diverse. I filtri sono del tipo passabanda: rispondono attivamente in un certo intervallo di frequenze spaziali. In altre parole, «vedono» solo variazioni di intensità da pixel a pixel che non sono né troppo rapide né troppo lente. Per ogni scala spaziale il processo di identificazione delle variazioni di intensità consiste pertanto nel filtrare l'immagine con un filtro (o campo recettivo) centro-periferia di una particolare dimensione e nel trovare i passaggi attraverso zero nell'immagine filtrata. Per una combinazione di scale è necessario disporre filtri di dimensioni diverse, che eseguano lo stesso calcolo per ciascuna scala. I filtri grandi pertanto scoprono contorni morbidi o sfumati nonché variazioni complessive di illuminazione; i filtri piccoli identificano particolari più dettagliati. I contorni netti verrebbero identificati a tutte le scale.

Risultati teorici rendono quest'idea veramente interessante, in quanto mostrano che caratteristiche simili ai passaggi attraverso zero in una immagine filtrata possono essere ricche di informazione. Ben Logan dei Bell Laboratories ha dimostrato che un segnale unidimensionale filtrato attraverso una certa classe di filtri può essere ricostruito dai suoi soli passaggi attraverso zero. Il laplaciano di una gaussiana non soddisfa esattamente le condizioni di Logan, tuttavia il suo teorema fa pensare che i simboli primitivi forniti dai passaggi attraverso zero costituiscano potenti simboli visivi. Più recentemente Alan Yuille e io abbiamo analizzato teoricamente il filtraggio centro-periferia e siamo riusciti a dimostrare che le mappe dei passaggi attraverso zero ottenute a scale diverse possono rappresentare l'immagine originale completamente, senza cioè perdita di informazione.

Ciò non vuol dire che i passaggi attraverso zero costituiscano l'unico schema di codificazione ottimale per un processo come la visione stereoscopica o l'unica base della visione biologica. Essi, o qualcosa di simile a essi, rappresentano tuttavia uno schema di codificazione che potrebbe rivelarsi ottimale e avere una posizione di rilievo tra gli elementi che debbono essere messi in corrispondenza tra le due immagini retiniche. Abbiamo dunque una possibile risposta alla domanda relativa a che cosa il modulo della visione stereoscopica metta in corrispondenza. Inoltre abbiamo l'inizio di una teoria computazionale che darà precisione matematica al vago concetto di «contorno» e lo collegherà a proprietà note della visione biologica, come la preminenza di cellule «rilevatrici di contorni» scoperta da

Il filtro biologico incorporato nelle cellule della retina assomiglia, per i suoi effetti, alla procedura per calcolatore mostrata nelle due pagine precedenti. Il filtro comincia con uno strato di fotorecettori, che misurano l'intensità della luce riflessa dal mondo visivo, e che sono collegati (attraverso cellule nervose intermedie non raffigurate nel diagramma) a uno strato di cellule gangliari retiniche, che trasmettono dati visivi ai centri visivi superiori del cervello. Qui è raffigurato solo un gruppo di collegamenti: un fotorecettore *(in rosso)* eccita una cellula gangliare «a centro ON» favorendo la sua tendenza a generare segnali nervosi; i fotorecettori della periferia *(in blu)* inibiscono la cellula gangliare. Questa disposizione equivale a un filtraggio biologico centro-periferia.

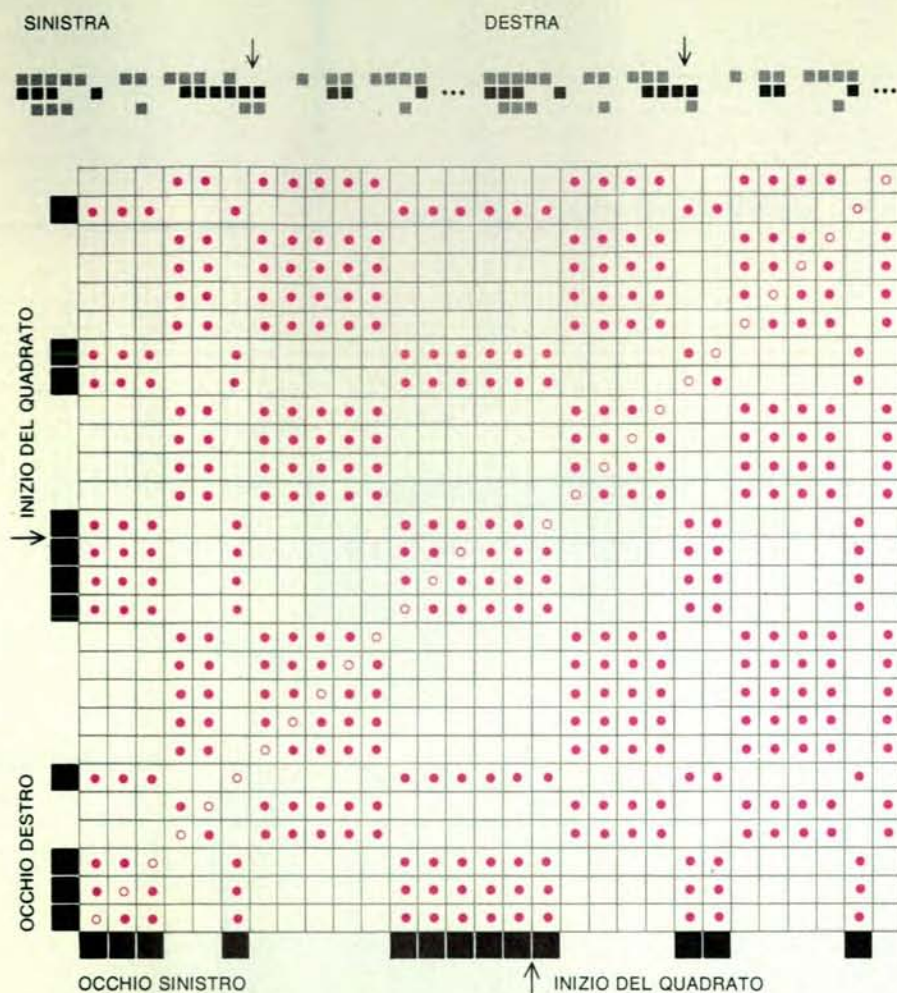
SINISTRA

DESTRA

INIZIO DEL QUADRATO

OCCHIO DESTRO

OCCHIO SINISTRO

INIZIO DEL QUADRATO

L'algoritmo della visione stereoscopica formulato da Marr e dall'autore ricostruisce il mondo visivo tridimensionale cercando corrispondenze fra punti su due righe in una coppia di stereogrammi a punti casuali. In alto nell'illustrazione si vedono due di queste righe *(in bianco e nero)*. Al di sotto, le righe sono collocate lungo gli assi di un diagramma. Le linee orizzontali nel diagramma rappresentano linee di vista per l'occhio destro; le linee verticali rappresentano linee di vista per l'occhio sinistro. I cerchietti pieni in colore indicano tutti i punti di intersezione nei quali ambedue gli occhi vedono un punto bianco o un punto nero. In linea di principio un punto nero in uno stereogramma potrebbe corrispondere a un qualunque punto nero nell'altro, e lo stesso vale per i punti bianchi. Tuttavia, solo alcune corrispondenze sono corrette *(cerchietti vuoti in colore)*, cioè solo alcune corrispondenze rivelano che un quadrato nella struttura a punti casuali ha una disparità binoculare. La spiegazione dell'algoritmo continua nell'illustrazione di pagina 80.

David H. Hubel e Torsten N. Wiesel della Harvard Medical School nella parte della corteccia cerebrale responsabile dell'elaborazione dell'informazione visiva.

Riassumendo, argomenti legati alla teoria della visione da una parte e dati biologici dall'altra fanno pensare che il rilevamento e l'identificazione di variazioni di intensità in una immagine a diverse scale spaziali costituiscano un primo importante passo per la visione stereoscopica e altri processi visivi. Un modo per realizzare tale passo è filtrare l'immagine con il laplaciano di una gaussiana; i passaggi attraverso zero nella matrice filtrata corrispondono ai contorni di intensità nell'immagine. Informazioni simili sono implicite nell'attività delle cellule gangliari a centro ON e di quelle a centro OFF della retina. Per rappresentare esplicitamente i passaggi attraverso zero (sempre ammesso che il cervello faccia qualcosa del genere) una classe di neuroni rilevatori di contorni nel cervello (probabilmente nella corteccia cerebrale) dovrebbe svolgere operazioni specifiche sull'uscita delle cellule a centro ON e delle cellule a centro OFF, che sono vicine nella retina. Qui ci scontriamo ancora una volta con la mancanza di informazioni sulle operazioni elementari che le cellule nervose sono in grado di eseguire. In tale campo di ricerca tuttavia V. Torre dell'Università di Genova e C. Koch e io del MIT abbiamo recentemente ottenuto molti risultati interessanti.

Vediamo ora di capire in che modo una rappresentazione delle variazioni di intensità possa essere utile per la visione stereoscopica. Consideriamo per cominciare un algoritmo formulato da Marr e da me, che realizza i vincoli di cui abbiamo discusso in precedenza, e cioè l'unicità (un punto dato su una superficie fisica ha una sola localizzazione, cosicché solo una corrispondenza binoculare è corretta) e la continuità (le variazioni di profondità in genere sono continue, cosicché le disparità binoculari tendono a variare con continuità). L'algoritmo ha successo nella risoluzione di stereogrammi a punti casuali di alcune immagini naturali. Viene eseguito da un calcolatore, pertanto la sua esecuzione effettiva non è altro che una successione di calcoli. Possiamo pensare tuttavia che l'algoritmo consista di una rete tridimensionale di nodi, in cui i nodi rappresentano tutte le possibili intersezioni di linee di vista che si dipartono dagli occhi nel mondo tridimensionale. Il vincolo dell'unicità si realizza imponendo che i nodi lungo una data linea di vista si inibiscano a vicenda. Al tempo stesso il vincolo della continuità si realizza imponendo che ogni nodo ecciti i suoi vicini. Nel caso degli stereogrammi a punti casuali la procedura è relativamente semplice: i corrispondenti dei pixel di ciascuna linea orizzontale di uno stereogramma vanno cercati solo nella corrispondente riga dell'altro stereogramma.

L'algoritmo inizia con l'assegnazione del valore 1 a tutti i nodi che rappresentano una corrispondenza binoculare fra due pixel bianchi e due pixel neri nella coppia di stereogrammi, mentre agli altri nodi viene assegnato il valore zero. Gli 1 pertanto contrassegnano tutte le corrispondenze, vere e false *(si veda l'illustrazione qui a fianco)*. Poi l'algoritmo esegue una somma algebrica per ciascun nodo. In questo calcolo i nodi vicini che hanno valore 1 contribuiscono con pesi positivi, mentre i nodi che hanno valore 1 lungo linee di vista contribuiscono con pesi negativi. Se il risultato supera un certo valore di soglia, al nodo si assegna valore 1; altrimenti il nodo è zero. Questo rappresenta una iterazione della procedura. Dopo qualche iterazione la rete raggiunge la stabilità. Il problema della visione stereoscopica è risolto.

L'algoritmo presenta alcuni grandi pregi. È un algoritmo cooperativo: consiste cioè di operazioni locali che potrebbero essere eseguite in modo asincrono e in parallelo da un gran numero di elaboratori molto semplici. Si pensa che i neuroni siano in grado di eseguire tali operazioni. Inoltre l'algoritmo può completare i dati mancanti: esso interpola superfici continue pur consentendo al tempo stesso discontinuità nette. La rete necessaria per elaborare immagini naturali ricche di particolari minuti dovrebbe tuttavia essere piuttosto ampia, e in ogni istante la maggior parte dei suoi nodi sarebbe inattiva. Inoltre i valori di intensità non sono soddisfacenti per immagini più naturali degli stereogrammi a punti casuali.

Si può estendere il dominio dell'algoritmo a immagini naturali filtrando dapprima le immagini con il laplaciano di una gaussiana per ottenere il segno della loro convoluzione. Le mappe binarie che ne risultano fungono da ingressi per l'algoritmo cooperativo. Le mappe stesse però danno da pensare. In quelle generate da grandi filtri a risoluzione spaziale corrispondentemente bassa, i passaggi attraverso zero di un dato segno (per esempio i

passaggi per i quali il segno della convoluzione cambia da positivo a negativo) risultano assari rari e non sono mai molto vicini. Quindi su una ampia gamma di disparità le false corrispondenze (corrispondenze tra passaggi attraverso zero non corrispondenti in una coppia di stereogrammi) sono essenzialmente assenti.

Questo fa pensare a una diversa classe di algoritmi per la visione stereoscopica. Un algoritmo del genere, sviluppato recentemente per robot da Nishihara, mette in corrispondenza chiazze positive o negative in coppie di immagini filtrate. Un altro algoritmo, sviluppato in precedenza da Marr e da me, mette in corrispondenza passaggi attraverso zero dello stesso segno in coppie di immagini realizzate per mezzo di filtri di tre o più dimensioni. In primo luogo vengono messe in corrispondenza le immagini filtrate a grandi scale, e si misurano le disparità binoculari; i risultati vengono utilizzati per mettere a registro le immagini a risoluzione più fine. Si potrebbero anche utilizzare caratteristiche monoculari come le tessiture. In seguito si applica un analogo processo di corrispondenza alle immagini filtrate con filtri di dimensioni intermedie. Infine il processo viene applicato alle immagini più finemente filtrate. A questo punto si conoscono le disparità binoculari nella coppia di stereogrammi, e così il problema della visione stereoscopica è ridotto a un problema di trigonometria.

W. Eric L. Grimson dell'Artificial Intelligence Laboratory ha messo a punto una estensione del nostro algoritmo e una sua realizzazione al calcolatore che funzionano assai bene per una applicazione tipica: l'analisi di fotografie aeree. Questa realizzazione simula molte delle proprietà della percezione umana della profondità: dà per esempio buoni risultati quando una delle stereoimmagini non è a fuoco. Tuttavia vi sono alcune sottili differenze. Ricerche recenti di John Mayhew e John P. Frisby all'Università di Sheffield e di Julesz ai Bell Laboratories dovrebbero chiarire la questione.

Le iterazioni dell'algoritmo (illustrate schematicamente) risolvono il problema della visione stereoscopica. L'algoritmo attribuisce valore 1 a tutte le intersezioni delle linee di vista contrassegnate da una corrispondenza, mentre alle altre attribuisce valore 0, e calcola una somma ponderata per ogni intersezione. Le superfici nel mondo tridimensionale tendono a variare con continuità in profondità: pertanto le intersezioni adiacenti con valore 1 contribuiscono con valori positivi alla somma. L'occhio vede solo una superficie lungo una data linea di vista; pertanto le intersezioni con valore 1 lungo linee di vista contribuiscono con valori negativi. Se il risultato supera un valore di soglia, l'intersezione assumerà il valore 1 nell'iterazione successiva; altrimenti verrà valutata 0. Dopo qualche iterazione il calcolo è completo: gli stereogrammi sono decodificati. In modo analogo si possono elaborare immagini naturali trasformate in matrici binarie (cioè in mappe a due valori di passaggi attraverso zero) dopo il filtraggio centro-periferia.

Che cosa si può dire della visione stereoscopica biologica? Gli algoritmi che ho descritto sono ancora lontani dal risolvere il problema della corrispondenza con l'efficienza che è propria del nostro cervello, ma ci suggeriscono come il problema possa venire risolto. Nel frattempo ricerche sulla corteccia cerebrale del gatto e del macaco hanno rivelato che alcuni neuroni nelle aree visive della corteccia segnalano disparità binoculari. Pochi mesi fa Gian F. Poggio della Johns Hopkins University School of Medicine ha scoperto neuroni corticali che segnalano la corretta disparità in stereogrammi a punti casuali in cui vi sono molte false corrispondenze. La sua scoperta, insieme con la nostra analisi computazionale della visione stereoscopica, promette di gettare nuova luce sui meccanismi cerebrali che sono alla base della percezione della profondità.

Da questo tipo di ricerca sulla visione dovrebbe emergere un aspetto caratteristico e importante dal punto di vista del metodo scientifico; esso riguarda il ruolo del cervello e del calcolatore nell'aiutarci reciprocamente a capire come l'uno e l'altro elaborino informazione. Da una parte abbiamo il calcolatore, strumento potente per provare teorie computazionali e algoritmi, e al tempo stesso guida nel disegno di esperimenti neurofisiologici e nella ricerca sul cervello. Dall'altra abbiamo lo studio del cervello da cui la scienza del calcolatore riceve impulsi fondamentali.

Alcuni studiosi hanno sostenuto che il cervello fornisce soltanto prove di esistenza, cioè costituisce la dimostrazione vivente del fatto che certi problemi abbiano una soluzione. In effetti essi si sbagliano, perché il cervello può fare molto di più, può cioè indicarci come cercare soluzioni. Il cervello è un elaboratore di informazione che si è evoluto nell'arco di milioni di anni al punto da eseguire taluni compiti con efficacia superlativa. Se lo consideriamo, con giustificata modestia, come uno strumento incerto della ragione, è solo perché tendiamo a essere più coscienti di quello che fa meno bene (aspetti recenti della nostra storia evolutiva, come la logica, la matematica, la filosofia) e tendiamo a non essere affatto coscienti dei suoi veri poteri, per esempio la capacità di vedere o il controllo motorio. In questi ultimi campi abbiamo molto da imparare dal cervello, ed è proprio in essi che dovremmo riuscire a misurare i nostri progressi nella scienza del calcolatore e nella robotica. Solo allora potremmo cominciare a vedere quanto sia ampio l'orizzonte che ci aspetta.

Le molecole che fanno aderire le cellule

Sono la base per la costruzione di un animale; compaiono durante lo sviluppo e, regolando movimenti cellulari e posizione dei tessuti, contribuiscono all'organogenesi e al modellamento dell'embrione

di Gerald M. Edelman

La crescita organizzata di un embrione a partire da un unico uovo fecondato fino a un organismo adulto funzionante ha sempre affascinato i biologi dai tempi antichi fino a oggi. Anche se gli eventi successivi dello sviluppo embrionale sono stati descritti in molte specie con minuzia di particolari, l'interrogativo fondamentale dell'embriologia lancia ancora una sfida alle più sofisticate risorse della biologia contemporanea: che cosa controlla il processo della morfogenesi, cioè dell'origine della forma? In altre parole, in che modo un codice genetico monodimensionale riesce alla fine a specificare un organismo tridimensionale?

Si deve poter rispondere a questo interrogativo a livello molecolare, ma è improbabile che la risposta venga per applicazione diretta dei metodi della biologia molecolare soltanto. I processi primari dello sviluppo - la divisione cellulare, il movimento cellulare, l'adesione tra cellule, il differenziamento cellulare e la morte cellulare - non possono venire analizzati in termini di liste di geni e di interazioni lineari dei loro prodotti proteici. Ciascuno di essi è il risultato di miriadi di eventi molecolari, regolati in parallelo non solo da geni che interagiscono ma anche da eventi concomitanti epigenetici, cioè non specificati direttamente da particolari geni. Data la natura microscopica di tutti questi eventi e la loro notevole complessità, non c'è forse da stupirsi che finora non esista alcuna adeguata teoria dello sviluppo nello stesso senso in cui vi sono valide teorie per la genetica e per l'evoluzione. La costruzione di una tale teoria rappresenta una sfida avvincente perché la determinazione della natura e della regolazione delle sequenze genetiche ed epigenetiche nel corso dello sviluppo embrionale e del loro rapporto con l'evoluzione è forse il massimo enigma ancora insoluto della biologia.

Una linea di condotta per cercare di risolvere parte del mistero consiste nello scegliere un solo processo primario, nel tentare di analizzarlo a livello molecolare e quindi di correlare tale analisi con quelle di altri processi primari, importanti per la morfogenesi. Quale processo di sviluppo si dovrebbe scegliere? La scelta dipende in qualche misura dal fatto che si sia interessati principalmente al differenziamento di singole cellule oppure alla formazione di un quadro più generale. Una strategia adottata nel nostro laboratorio alla Rockefeller University, dove prevale l'interesse per la seconda alternativa, consiste nel mettere bene a fuoco il processo primario per mezzo del quale le cellule aderiscono l'una all'altra. Questa adesione per stabilire la forma deve verificarsi a ogni stadio dello sviluppo ed è l'evento più suscettibile all'attacco diretto da parte di sostanze chimiche; inoltre è il processo che in maniera più ovvia è in relazione con il mantenimento della forma finale. La sua analisi è particolarmente interessante perché essa è mediata da molecole proteiche.

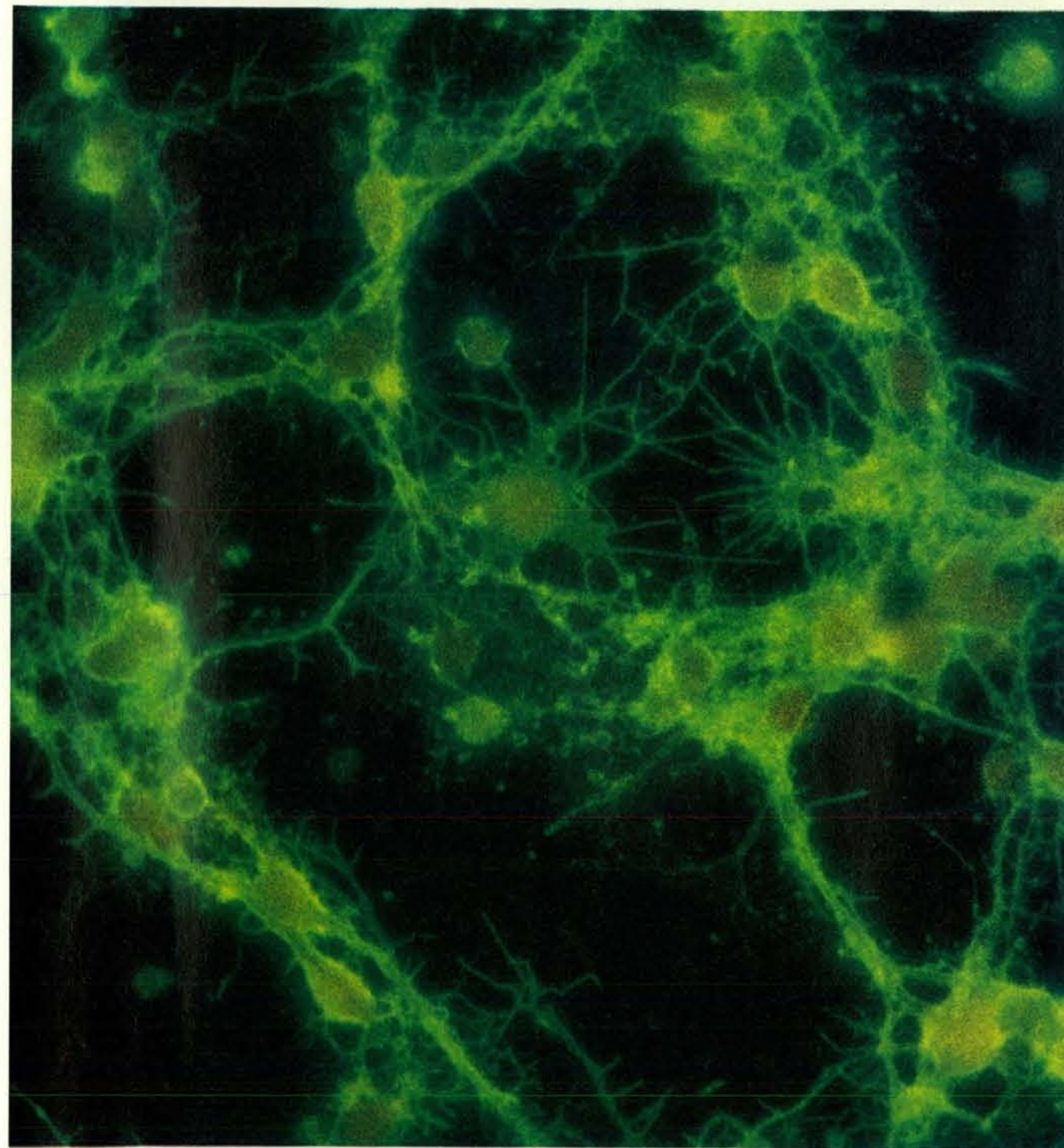
Le proteine sono specificate direttamente dal codice genetico e pertanto sia la loro funzione nel determinare la forma dell'embrione sia il controllo sui geni che ne sono responsabili si possono collegare all'interno di un unico sistema. Anche se si potrà capire completamente la determinazione della forma dell'embrione solo quando saranno stati analizzati tutti i processi primari, le ricerche compiute sull'adesione tra cellule nell'ultimo decennio forniscono un incoraggiante avvio, dato che lasciano intravvedere ipotesi molto suggestive su come potrebbero alla fine essere risolti taluni aspetti dell'enigma della morfogenesi.

Per dare un'idea della natura del problema, mi si lasci brevemente passare in rassegna alcuni importanti eventi precoci dello sviluppo dell'embrione di pollo. Poco dopo la fecondazione, una successione di divisioni cellulari dà luogo alla formazione del blastoderma, un disco piatto sulla membrana che avvolge l'uovo, al di sopra del tuorlo. In esso le cellule più superficiali costituiscono l'epiblasto, il quale dà origine a una lamina sottostante, l'ipoblasto e, più tardi, a uno strato intermedio, il mesoblasto.

Nella morfogenesi precoce l'evento fondamentale è costituito dalla formazione, a partire da questi precursori cellulari, dei tre foglietti embrionali: l'ectoderma, il mesoderma e l'endoderma. In questo processo, detto gastrulazione, i movimenti cellulari e le alterazioni delle varie lamine di tessuto, che seguono tutti uno schema ben ordinato, si susseguono in una ben precisa sequenza. L'epiblasto dà origine all'ectoderma e, assieme a parti dell'ipoblasto, all'endoderma; il mesoblasto dà origine al mesoderma. In stadi più tardivi, l'ectoderma, che è lo strato più superficiale dell'embrione, si sviluppa formando il sistema nervoso e l'epidermide. L'endoderma dà origine ai polmoni, all'intestino e ad altri organi. Il mesoderma, infine, produce i muscoli, le ossa e i reni.

Il processo che ha un'importanza chiave per tutti i successivi eventi è l'induzione embrionale, nella quale le cellule che si differenziano per dare origine a certe strutture in un determinato strato richiedono la presenza di altre cellule caratterizzate da una storia diversa (e appartenenti allo stesso strato o a uno strato diverso). L'apposizione di cellule si realizza di norma mediante movimenti morfogenetici.

L'induzione, o differenziamento che dipende dall'ambiente, ha inizio con la gastrulazione. Nell'embrione essa è il risultato di sequenze ininterrotte di interazioni reciproche tra differenti gruppi di cellule e si effettua o per contatto cellulare diretto o per trasferimento di fattori molecolari. Tra tutte queste induzioni la più precoce e la più critica è quella che fa diventare le cellule ectodermiche precursori di tessuti nervosi come l'encefalo e il midollo spinale. Quest'induzione primaria o neurale dipende da interazioni che si stabiliscono tra il mesoderma e l'ectoderma e dà luogo alla formazione della placca neurale dall'ectoderma. In seguito, la placca neurale va incontro a una serie coordinata di ripiegature (neurulazione) per formare il tubo neurale. Parti diverse del sottostante mesoderma hanno effetti differenti nel determinare il prosencefalo, il mesencefalo, il metencefalo e il midollo spinale. L'induzione presenta una specializzazione regionale perché è il risultato di una sequenza nel tempo e nello spazio, all'interno dell'embrione che via via si modifica, di segnali di stimolazione o di inibizione.

Precocemente, nel corso dell'embriogenesi, alcuni eventi morfogenetici che riflettono i vari processi primari sono realmente visibili: così lo sfrecciare di cellule endodermiche verso l'interno del blastoderma durante la gastrulazione, passando attraverso quel solco che è detto linea primitiva; la segregazione di cellule

In questa microfotografia, ottenuta al microscopio a fluorescenza da Cheng-Ming Chuong nel laboratorio dell'autore, la N-CAM, una molecola che favorisce l'adesione tra cellule e che è associata con il tessuto nervoso, appare sui corpi cellulari e sui processi in fase di allungamento dei neuroni di pollo coltivati in vitro, che qui sono ingranditi all'incirca di 1200 diametri. Le cellule in questione sono state trattate con un anticorpo contro la N-CAM, marcato con la fluoresceina, un colorante fluorescente che, illuminato con radiazioni ultraviolette, emette una luminescenza verde. L'anticorpo si è legato alle CAM su tutte le superfici cellulari, inclusi i coni di crescita, strutture mobili a spatola, alle estremità dei processi neuronali in fase di allungamento, di cui parecchie risultano bene evidenti al centro della microfotografia.

mesodermiche per formare i somiti, che sono i precursori segmentali sia di strutture muscolari sia di strutture scheletriche; il raggrupparsi di cellule ectodermiche che si allungano a dare quelle strutture che sono note come placodi, strutture che, alla fine, daranno origine a derivati specializzati come i gangli nervosi e il cristallino dell'occhio. I complessi organi di origine endodermica, quali il fegato e il pancreas, si formano dall'endoderma dell'intestino primitivo. Gli organi di origine mesodermica come il rene, con le sue strutture tubulari complesse, si formano dopo una serie di induzioni reciproche.

Un processo di particolare importanza perché si realizzino le giuste apposizioni e induzioni di tessuti è costituito dalla conversione di strati di cellule attaccate le une alle altre (epiteli) in masse migranti formate da cellule sciolte, vagamente organizzate (mesenchima); il processo inverso è altrettanto importante in certe regioni dell'embrione.

Lo sviluppo ha un suo decorso storico. Non solo l'induzione primaria ma l'intera serie di induzioni secondarie, che danno gli organi che sono stati menzionati prima, dipende dall'apposizione di particolari cellule che hanno storie diverse. Questa straordinaria successione di eventi si svolge in una sequenza temporale ben definita e si organizza in stadi, in ognuno dei quali vi sono strutture che servono come base di strutture che compariranno in stadi successivi.

In che modo le descrizioni classiche della morfogenesi possono venir tradotte nel linguaggio delle molecole? Come ho indicato, la nostra strategia è consistita nel ricercare molecole che mediassero l'adesione tra cellule allo scopo di studiarne la struttura e il modo in cui operano e quindi per correlare i nostri risultati agli eventi e alle sequenze chiave che si svolgono in altri processi primari di sviluppo. Prima di descrivere i risultati sperimentali analizzerò alcuni concetti generali che correlano il fenomeno dell'adesione a quello di configurazione, perché la formazione di una certa configurazione costituisce il tema fondamentale della biologia dello sviluppo.

A livello cellulare vi sono due modi in alternativa per mezzo dei quali potrebbero formarsi delle configurazioni senza ricorrere all'intervento diretto di un qualche «piccolo architetto» o «demone della costruzione». Il primo modo esigerebbe di contrassegnare preventivamente tutte le cellule con vari marcatori molecolari (presumibilmente proteine), ciascuno con una complementarità spaziale nei riguardi di qualche altro marcatore presente su una cellula che verrà a porsi vicino alla prima cellula nella configurazione. Questo è essenzialmente il modo in cui sono state costruite parti della grande abbazia di Mont-Saint-Michel, vicino alla costa della Normandia. Le pietre furono tagliate e modellate sulla terraferma, contrassegnate dai costruttori e trasportate poi sull'isola secondo un piano. Questo modello del Mont-Saint-Michel è una metafora per varie teorie sull'adesione tra cellule, basate sulla «chemoaffinità». La principale difficoltà che emerge da tali teorie è che, se la configurazione che si deve formare è complessa, se presenta una notevole variazione nella forma, o ha molti elementi e un notevole dettaglio locale (come, per esempio, nell'encefalo), il numero di marcatori superficiali specifici che determinano la localizzazione di ogni cellula deve essere enorme. Dato che poi questi marcatori devono essere con tutta probabilità proteine specifiche, ciascuna codificata da un differente gene, il numero di geni sarebbe corrispondentemente elevato. Come nel caso di una chiave di volta di un arco a cui sia stato assegnato un numero sbagliato, così un marcatore cellulare che si venga a trovare in un posto sbagliato provocherebbe un disastro. Va detto, inoltre, che una configurazione siffatta è prefigurata e essenzialmente statica: una volta che i giusti marcatori sono stati riuniti, non è più necessario alcun dinamismo.

Vi è un altro modo più dinamico di produrre configurazioni, un modo analogo a ciò che si potrebbe osservare in un corso d'acqua di montagna. In una simile situazione cinetica, lontana dall'equilibrio, la configurazione che emerge dipende dal ruolo dell'energia che viene dissipata nell'ambiente e che deve opporsi a varie limitazioni. Per prendere il caso più semplice in questo modello, si immagini un corso d'acqua che scenda da un fianco della montagna e vada a colpire un masso roccioso sommerso, la cui temperatura sia al disotto del punto di congelamento dell'acqua. All'inizio il flusso dell'acqua sarà solo leggermente influenzato dal masso e il corso rimarrà unico. Con il tempo, tuttavia, a mano a mano che l'acqua a contatto con il masso si congela, questa struttura si fa più voluminosa e improvvisamente potrà diventare una barriera, che suddividerà il corso in due facendogli assumere una nuova forma mentre scende dalla montagna. Tutti i modellamenti successivi saranno influenzati dall'effetto di quel congelamento iniziale. A valle, i rivoletti potranno frammentarsi in una varietà di nuove e intricate figure allorché, ai livelli più bassi, incontreranno differenti ostacoli. Visto dall'alto, l'intero corso avrà nondimeno una conformazione ben definita.

Naturalmente, in un embrione le cose sono più complicate. Ciononostante si può immaginare che i corrispondenti cellulari della forza motrice del potenziale gravitazionale sulla montagna siano i processi della divisione cellulare, della morte cellulare e del movimento cellulare. Il corrispondente del congelamento dell'acqua sarebbe l'adesione delle cellule l'una all'altra per mezzo di apposite molecole dette CAM (da *Cell-Adhesion Molecules*), distribuite sulla superficie cellulare. Il congelamento completo corrisponderebbe a un legame tanto rigido tra le cellule che certune non potrebbero più muoversi l'una rispetto all'altra. La cosa importante da notare è che, per ottenere un dato numero di configurazioni cellulari in un processo analogo al modello del corso d'acqua di montagna, il numero di differenti CAM specifiche non deve essere necessariamente grande quasi come quello dei contrassegni di riconoscimento, postulati nel modello del Mont-Saint-Michel, purché soltanto vi sia qualche mezzo dinamico per modificare l'intensità del legame tra le varie CAM. Quasi certamente una sola di queste molecole non sarà sufficiente; pur con la gamma dinamica del modello cinetico, alcune molecole con nuove specificità potranno essere necessarie per generare determinate configurazioni.

Questa discussione dovrebbe rendere chiara la necessità di distinguere l'adesione tra cellule dal riconoscimento cellulare. Se è interessata la prossimità delle cellule l'una all'altra, il riconoscimento mediante marcatori prefigurati sottintende chiaramente una specificità locale (come nel modello del Mont-Saint-Michel), mentre, se l'adesione tra le varie cellule è soggetta a cambiamenti dinamici, essa non richiede in sé un simile grado di riconoscimento. La configurazione può anche formarsi senza il riconoscimento individuale tra cellule e, pertanto, si deve distinguere tra specificità di riconoscimento e selettività (o fissazione della configurazione) che può emergere dall'interazione tra diversi processi dinamici.

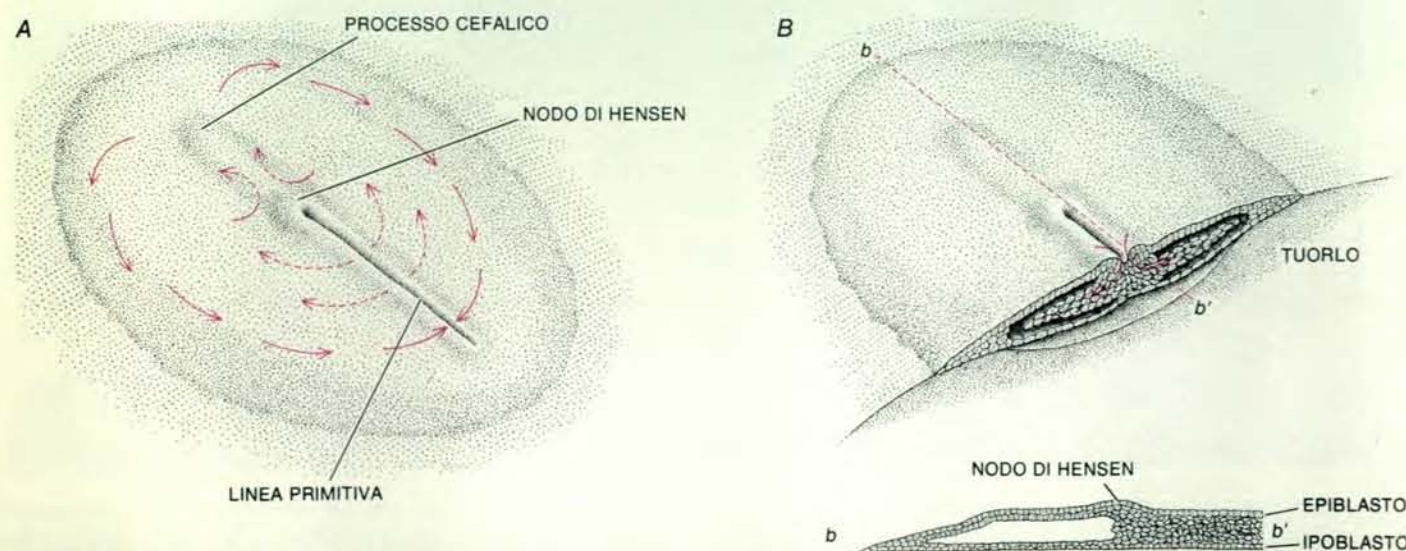
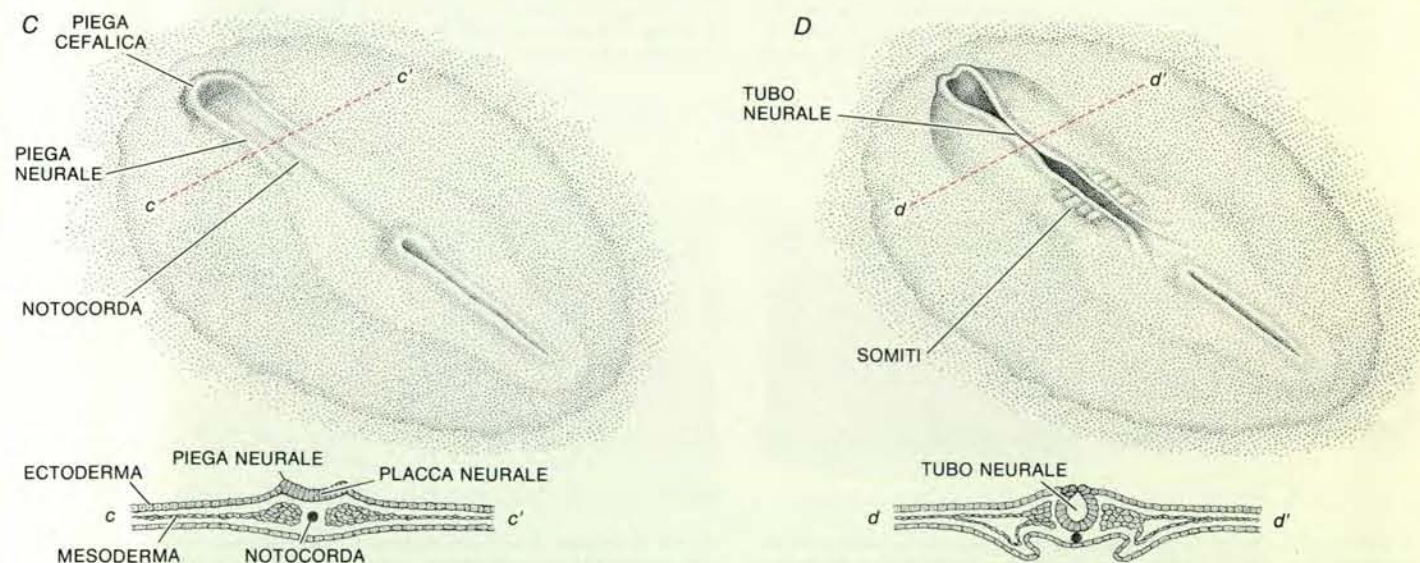
Un'altra distinzione ancora deve essere fatta. I nomi delle differenti cellule che si legano assieme non dovrebbero essere confusi con i nomi delle differenti molecole che mediano il loro legame. Per convenzione chiamiamo il legame tra cellule simili «omotipico» e il legame tra molecole simili «omofilo». La distinzione non è banale: si può avere un legame eterotipico (tra differenti tipi di cellule) mediato da un meccanismo omofilo (dovuto a molecole simili).

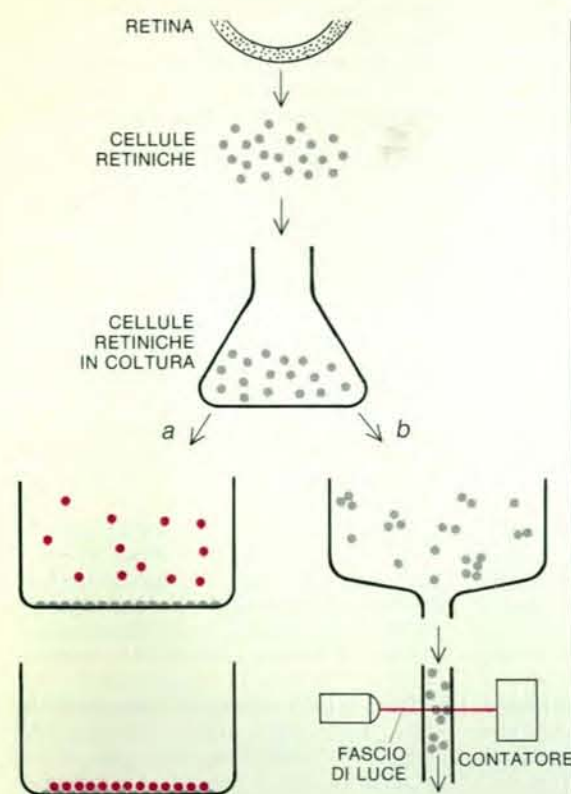
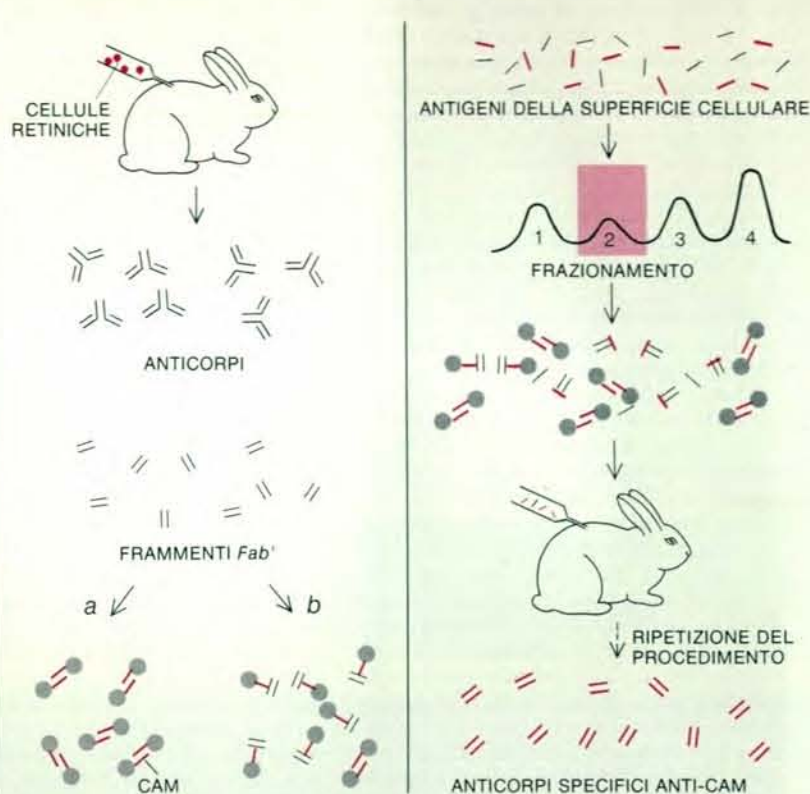
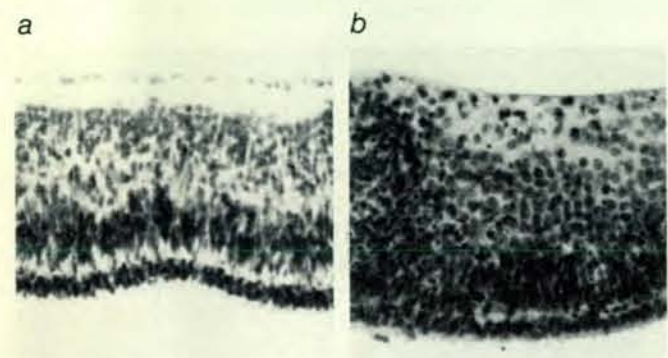
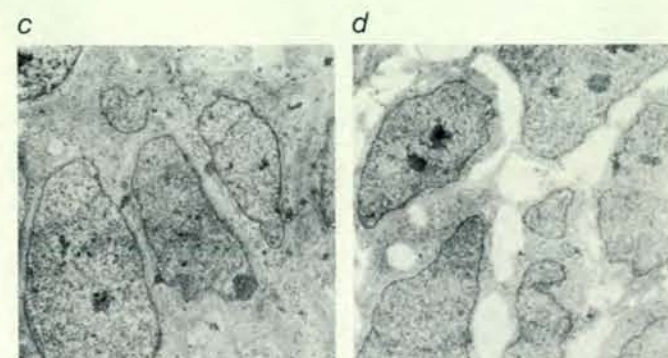
I miei collaboratori e io abbiamo cominciato la nostra ricerca sulle CAM partendo dal presupposto che dovesse trattarsi di molecole presenti sulla superficie cellulare, probabilmente proteine e probabilmente anche scarse. Un modo per individuare una proteina scarsa che abbia un'attività ben definita consiste nel preparare anticorpi specifici in grado di inibirne l'attività. L'antigene purificato che si lega a tali anticorpi è probabilmente la proteina desiderata. Nel processo di purificazione, a ogni passo è necessario effettuare saggi che misurino di quanto viene aumentata l'attività (nel caso di ipotetiche CAM, l'adesione tra le cellule) a mano a mano che la proteina viene purificata. Abbiamo messo a punto due tipi di saggi: saggi di identificazione, che rivelerebbero la presenza di CAM e ne misurerebbero l'attività, e saggi di perturbazione, che rivelerebbero la capacità di anticorpi specifici contro ipotetiche CAM di interferire con l'adesione tra cellule, che le CAM mediano, e quindi di demolire determinate configurazioni tissutali.

Per realizzare i saggi di identificazione, abbiamo dissociato le cellule di un tessuto digerendolo con l'enzima tripsina, che ha scisso le molecole proteiche presenti sulla superficie delle cellule. Dopo un periodo di ricupero, durante il quale le cellule isolate potevano nuovamente sintetizzare le loro proteine superficiali, queste cellule sono state lasciate collidere e aderire l'una all'altra.

Per identificare le molecole specifiche responsabili del legame tra superfici cellulari, abbiamo preso spunto da Günther Gerisch del Max Planck Institut für Biochemie, il quale, per perturbare l'adesione nei missomiceti, ha utilizzato degli anticorpi, e abbiamo cercato anticorpi contro le CAM che bloccassero l'adesione in maniera specifica. Abbiamo immunizzato conigli con tessuto cerebrale e retinico di pollo e abbiamo analizzato nei risultanti anticorpi la capacità di bloccare l'adesione tra cellule. (Gli anticorpi a forma di *Y* sono stati scissi per produrre frammenti «univalenti», detti frammenti *Fab'*, ciascuno dei quali con un unico sito in grado di legarsi a una particolare molecola antigenica. Gli anticorpi non scissi sono bivalenti e, invece di bloccare l'adesione, legherebbero due cellule assieme mediante le CAM.) Dopo un anno di laboriose ricerche, i nostri saggi di identificazione hanno rivelato la presenza di anticorpi specifici in grado di bloccare l'adesione tra cellule. Tuttavia, nei sieri di coniglio, questi anticorpi inibitori erano mescolati con altri anticorpi diretti non contro una CAM, ma contro altre molecole della superficie cellulare; a causa della loro eterogeneità, essi non potevano servire da sonde specifiche per identificare le CAM o per isolarle.

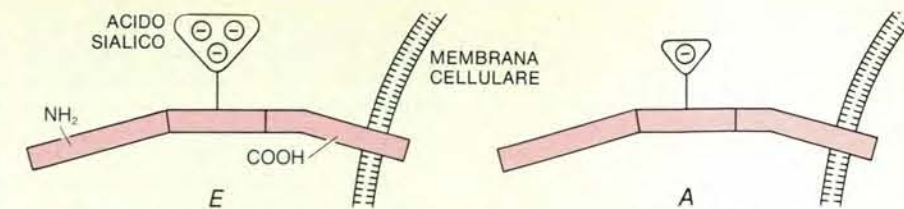
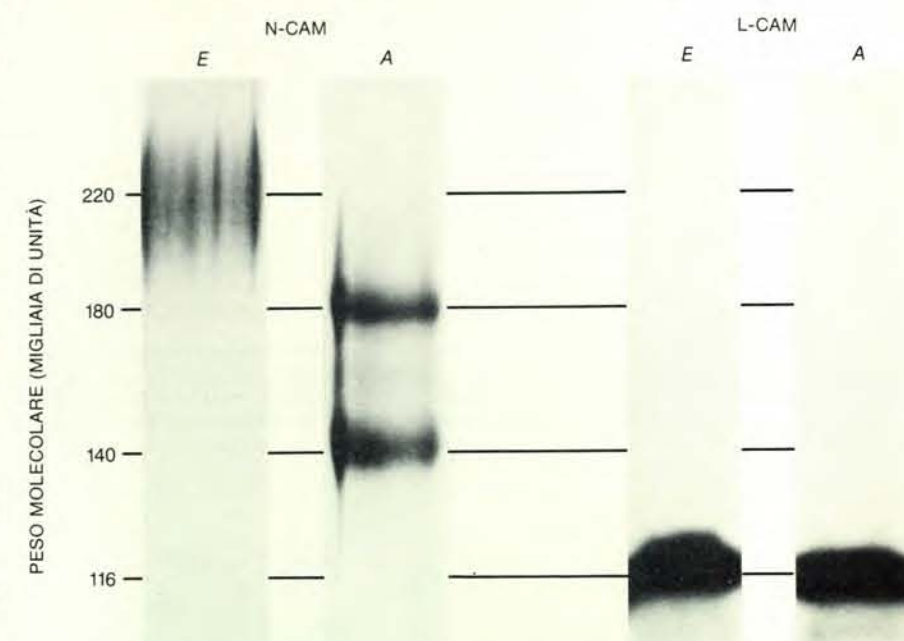
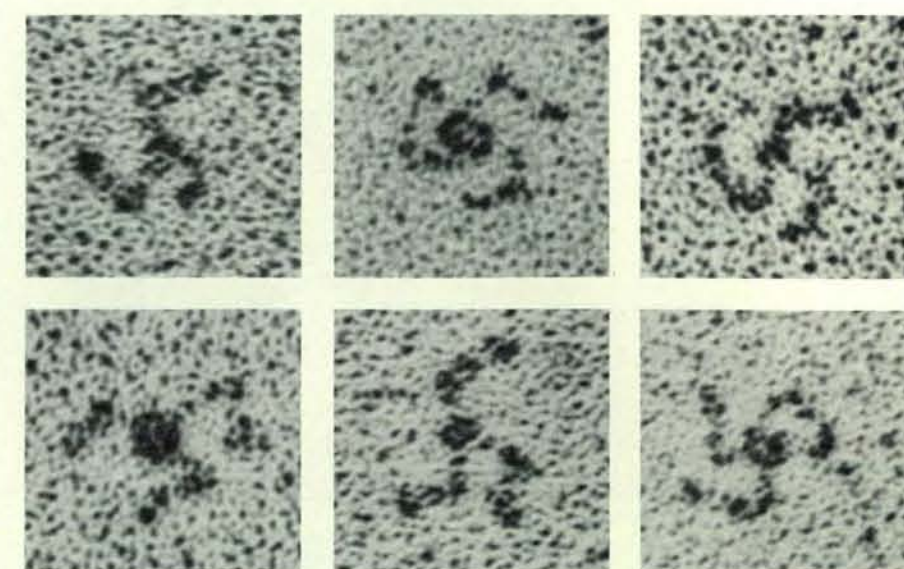
Pertanto abbiamo ideato un saggio di «neutralizzazione» per ridurre l'eterogeneità della popolazione degli anticorpi. Frazioni successive di antigeni di superficie sono state saggiate per valutare la loro capacità di legarsi agli anticorpi responsabili dell'inibizione, neutralizzando così la loro attività inibitoria. Sono stati di nuovo immunizzati conigli con le frazioni di antigeni di superficie relativamente arricchite di CAM e in grado di procedere alla neutralizzazione.

L'iterazione del suddetto procedimento ha generato frazioni antigeniche abbastanza pure da stimolare nei conigli la formazione di anticorpi anti-CAM fortemente specifici e nei topi anticorpi monoclonali ancor più specifici. Una vol-

Nello sviluppo precoce dell'embrione di pollo gli eventi morfogenetici chiave sono la gastrulazione *(A, B)* e la neurulazione *(C, D)*. Nella parte finale della gastrulazione le cellule dello strato superficiale del blastoderma convergono verso il solco centrale *(frecce continue)*, o linea primitiva. I movimenti delle cellule del mesoblasto (strato intermedio) sono indicati dalle frecce tratteggiate *(A)*. Una sezione trasversale *(B)* mostra come queste cellule si disperdano e formino il futuro mesoblasto e, infine, il mesoderma. La sezione sagittale *(b–b')* è condotta lungo la linea primitiva. Nell'induzione primaria (neurale), le interazioni tra mesoderma ed endoderma danno luogo alla placca neurale *(C)* che, ripiegandosi, forma il tubo neurale; subito dopo le cellule mesodermiche si separano e danno luogo ai somiti *(D)*.

ta conosciuta la natura degli antigeni CAM, siamo riusciti a produrre da numerose specie animali anticorpi specifici contro di essi.

Con una tale varietà di anticorpi in mano, abbiamo potuto controllare la distribuzione delle CAM in differenti tessuti e anche cercato in maniera specifica di sconvolgere lo sviluppo di configurazioni tissutali. Per esempio, marcando gli anticorpi (indotti da antigeni provenienti dall'encefalo e dalla retina) con una sostanza fluorescente, siamo riusciti ben presto a mostrare che una particolare molecola per l'adesione di cellule nervose, che abbiamo designato come N-CAM, è presente su tutti i neuroni sia nel sistema nervoso centrale sia in quello periferico. Gli anticorpi contro le N-CAM provocavano uno sconvolgimento nell'ordinato sviluppo dei tessuti nervosi, per esempio quelli della retina, allorché si sviluppavano in vitro. Tecniche analoghe hanno condotto all'identificazione di due altre molecole per l'adesione tra cellule: la L-CAM, originariamente isolata intatta da cellule epatiche embrionali da Bruce A. Cunningham e Warren J. Gallin, e la Ng-CAM (Ng sta per neurone-glia), isolata dai neuroni da Martin Grumet e da me.

Le CAM vengono isolate con questi procedimenti che sono stati messi a punto dall'autore in collaborazione con Robert Brackenbury, Urs Rutishauser e Jean-Paul Thiery. Prove di adesione *(a sinistra)* saggiano, a ogni stadio, l'attività delle CAM. Le cellule retiniche di pollo vengono dissociate e poi viene fatto in modo che risintetizzino le molecole che portano in superficie. L'adesione viene saggiata o conteggiando le cellule marcate che si legano a uno strato ben preciso di cellule *(a)*, o rimescolando le cellule in un agitatore e misurando poi in quale misura esse si aggregano *(b)*. In un saggio di inibizione *(al centro)*, conigli immunizzati con cellule retiniche di pollo producono anticorpi contro le proteine delle superfici cellulari. Questi anticorpi vengono scissi e danno frammenti *Fab'* univalenti. Le cellule retiniche di pollo di solito aderiscono l'una all'altra *(a)*, mentre i frammenti di anticorpi anti-CAM inibiscono questa adesione legandosi alle CAM *(in colore)*, che mediano l'adesione *(b)*. Un saggio di neutralizzazione *(a destra)* è necessario per identificare anticorpi specifici anti-CAM. Gli antigeni di superficie (comprese le CAM), ottenuti dalle cellule retiniche, vengono frazionati e le varie frazioni sono sottoposte a una prova per valutare la loro capacità di legarsi agli anticorpi responsabili dell'inibizione e quindi di neutralizzare l'inibizione. Una frazione *(in colore)* che neutralizza efficacemente viene quindi identificata in quanto ricca di CAM; pertanto il siero di sangue dei conigli immunizzati con una simile frazione risultano arricchiti di anticorpi anti-CAM. L'iterazione del procedimento genera anticorpi anti-CAM che sono molto specifici.

L'anticorpo contro la N-CAM perturba lo sviluppo del tessuto nervoso dell'embrione. Sono state rimosse retine da embrioni di pollo di sei giorni e coltivate per tre giorni con anticorpi normali di coniglio (frammenti *Fab'* univalenti), usati come controllo *(a, c)*, e con anticorpi anti-N-CAM estremamente specifici *(b, d)*. Le microfotografie mostrano che la struttura stratificata della retina *(a)* è alterata dagli anticorpi anti-N-CAM *(b)*. Quelle elettroniche, in cui compaiono cellule degli stessi tessuti, contrappongono la stretta apposizione delle membrane cellulari nel campione normale *(c)* con il meno esteso contatto cellulare e le ampie zone di spazio extracellulare del tessuto retinico alterato *(d)*.

Mediante un frazionamento cromatografico classico e cromatografia per affinità (che dipende dalla specificità degli anticorpi per le CAM), siamo riusciti a purificare la N-CAM e la L-CAM fino al punto in cui Cunningham e collaboratori hanno potuto cominciare ad analizzarne la struttura. Ambedue le molecole sono glicoproteine, cioè proteine alle quali sono attaccati gruppi di molecole glucidiche. La N-CAM è insolita per il fatto che contiene quantitativi straordinariamente grandi di uno zucchero complesso con carica negativa, chiamato acido sialico, il quale si presenta in una forma insolita polimerizzata.

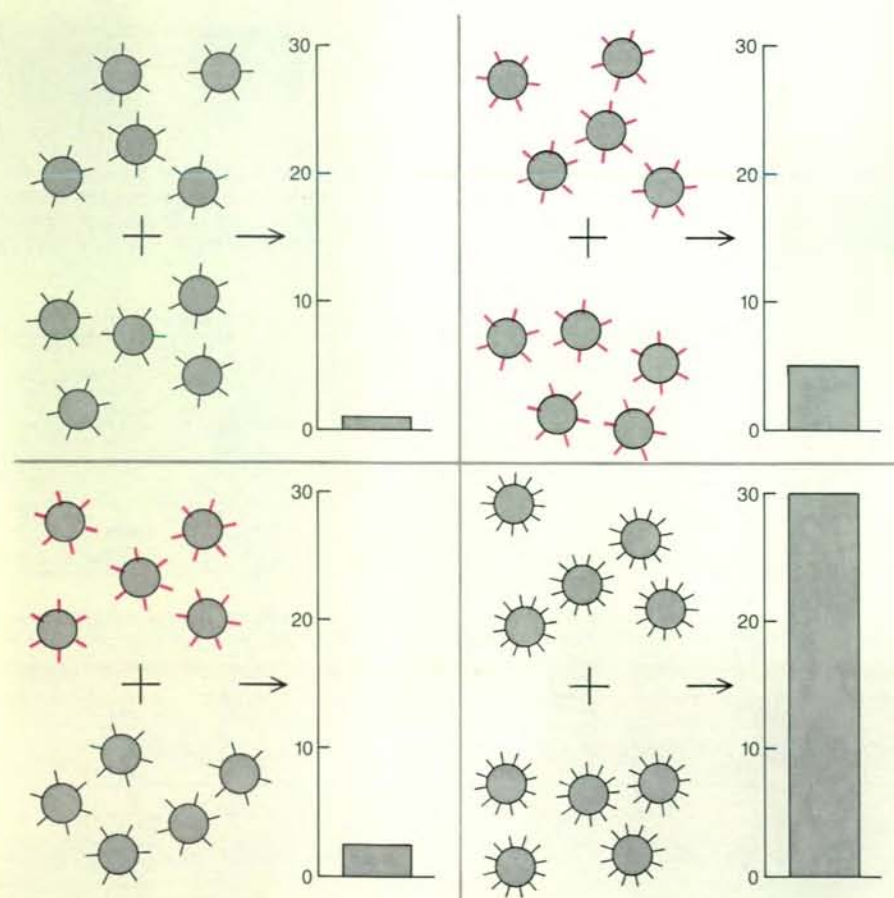
Una volta purificate le molecole per l'adesione tra le cellule, abbiamo potuto scinderle per via enzimatica così da costruire mappe lineari di N-CAM e L-CAM. Queste mappe hanno messo in evidenza che tali molecole sono strutturalmente diverse. La catena proteica dell'N-CAM (che, escluso lo zucchero, ha un peso molecolare massimo di 160000) consiste di tre aree legate da due segmenti di catena che sono particolarmente suscettibili all'azione litica degli enzimi. L'area con il gruppo amminico terminale (N-terminale) sporge dalla cellula e incorpora una regione che si lega a una regione corrispondente di una N-CAM, situata su un'altra cellula. Un'area intermedia contiene la maggior parte dell'acido sialico. L'area con il gruppo carbossilico terminale (C-terminale), infine, è associata con la membrana cellulare e parte di essa sembra inserita nel doppio strato di molecole lipidiche che costituiscono quella membrana.

Abbiamo trovato che la catena completa della N-CAM poteva facilmente associarsi con vescicole lipidiche artificiali, così da produrre un buon sistema sperimentale per lo studio del comportamento di legame della molecola. Per esempio, le vescicole che hanno le N-CAM si legano alle cellule nervose, ma se queste cellule vengono prima trattate con frammenti di anticorpi anti-N-CAM, le vescicole non si legano più. Questo ha fatto da supporto alla nostra iniziale congettura che il meccanismo di adesione sia omofilo, cioè che una N-CAM presente su una determinata cellula si leghi con una N-CAM che si trova su un'altra cellula.

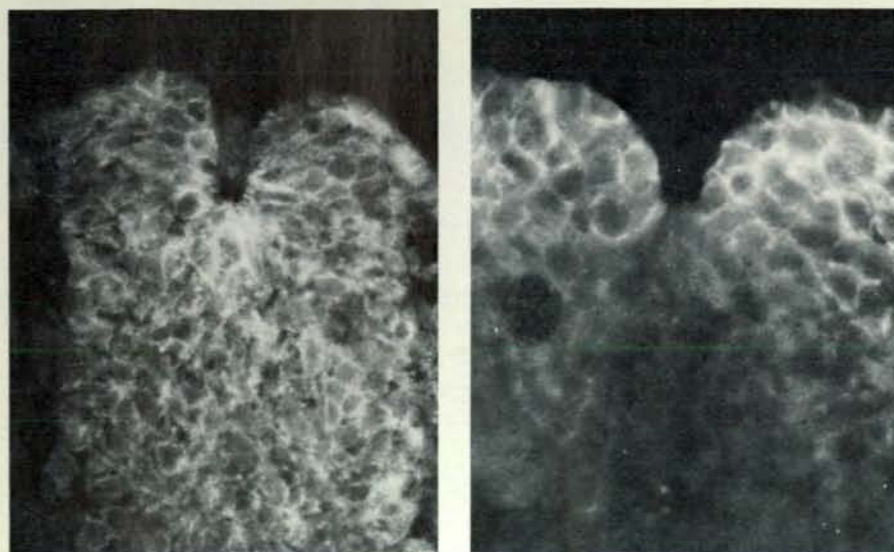
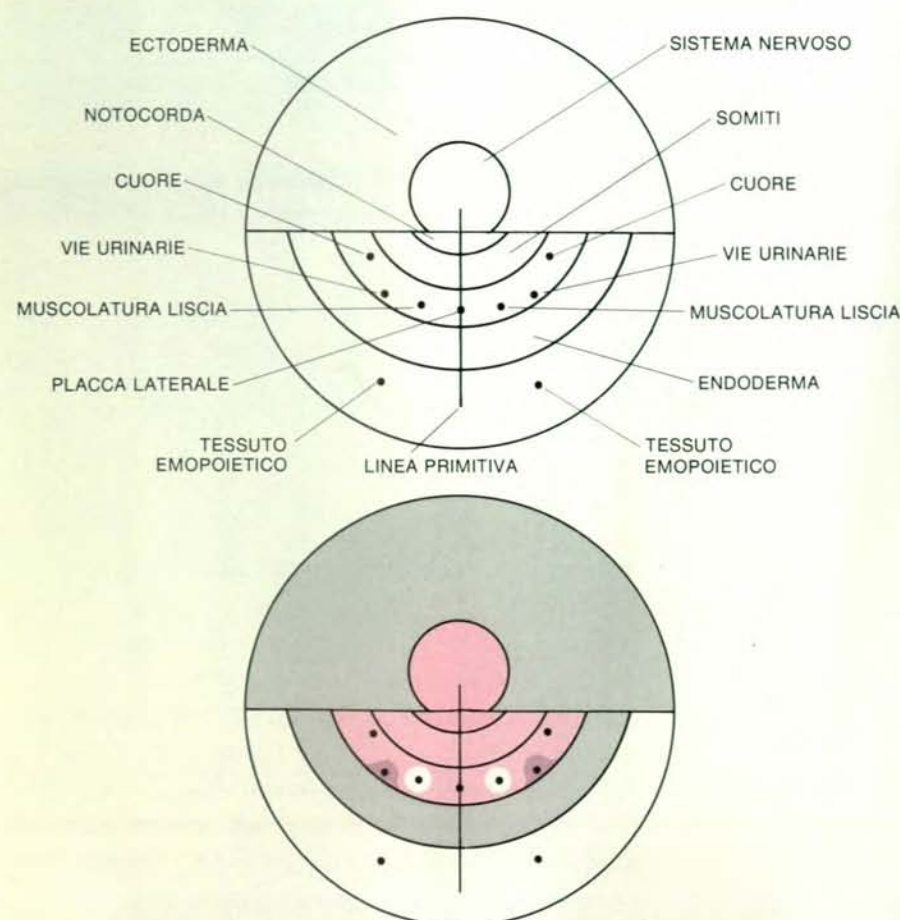
L'ipotesi che nelle N-CAM possano esservi diverse aree ha ricevuto sostegno di recente da studi preliminari di microscopia elettronica, effettuati in collaborazione con Carolyn Cohen della Brandeis University. Quando una N-CAM viene montata su una griglia e ombreggiata con metallo, alcune molecole (ma non tutte) esibiscono una notevole struttura triparti-

La N-CAM, molecola per l'adesione tra cellule nervose, è qui schematizzata nella sua forma embrionale *(E)*, a sinistra, e nella sua forma adulta *(A)*, a destra. La catena proteica *(in colore)* consta di tre aree: una amminoterminale o N-terminale *(NH_2)*, che presenta una regione di legame; una intermedia, che contiene gruppi di molecole di acido sialico, con carica negativa, e una zona carbossilicoterminale o C-terminale *(COOH)* che è associata con la membrana cellulare (e probabilmente inserita in essa). La forma *E* ha una quantità di acido sialico tre volte superiore.

Le differenze nel tasso di acido sialico sono rilevate dall'elettroforesi. La N-CAM di embrione di pollo *(E)* dà una banda diffusa con peso molecolare tra 200000 e 250000, mentre la N-CAM di pollo adulto *(A)* dà due bande più leggere *(a sinistra)*. In un'altra molecola per l'adesione tra cellule, la L-CAM, non si notano differenze elettroforetiche tra *E* e *A (a destra)*.

Singole N-CAM purificate da Bruce A. Cunningham compaiono in queste microfotografie elettroniche di Carolyn Cohen della Brandeis University. Alcune, montate su una griglia per microscopio elettronico e ombreggiate con metallo, formano strutture tripartite, i «trisceli», di cui se ne vedono qui sei, ingrandite di 440000 diametri. Ogni braccio del triscele può essere la catena proteica di una N-CAM. Non si sa se questa assuma tale forma sulla superficie della cellula.

ta, detta «triscele». Presumibilmente ciascuno dei tre bracci della struttura corrisponde a una catena proteica della N-CAM. Ogni braccio appare piegato in regioni che possono corrispondere a singole aree. Ulteriori studi effettuati con le N-CAM in membrane possono rivelare altre forme della molecola e mostrare se mai una qualsiasi delle conformazioni delle N-CAM nella cellula vivente assomigli alle conformazioni osservate al microscopio elettronico.

Ci siamo chiesti se, per caso, l'insolito contenuto di acido sialico delle CAM avesse un ruolo nel legame tra cellula e cellula. La risposta è stata che non ha alcun ruolo diretto: le molecole di N-CAM dalle quali è stato rimosso tutto l'acido sialico dall'enzima sialidasi si legheranno ancora in modo specifico alle cellule. Questo risultato, però, non ha escluso un ruolo indiretto dell'acido sialico nel legame. In effetti, due osservazioni hanno mostrato una funzione inattesa, ma molto importante, per questo zucchero. La prima osservazione è stata che vi sono due forme principali della molecola di N-CAM. La N-CAM di encefalo embrionale ha 30 grammi di acido sialico ogni 100 grammi di proteina. Quando viene sottoposta a elettroforesi migra sotto forma di una banda diffusa con un peso molecolare che varia tra 200 000 e 250 000. Per contro, la N-CAM da encefalo adulto ha solo 10 grammi di acido sialico ogni 100 grammi di proteina e forma sul gel elettroforetico due o tre bande ben definite. In un certo momento durante lo sviluppo, la forma embrionale *(E)* della molecola deve essere convertita in una delle numerose forme adulte *(A)*, o scambiata con essa. La seconda osservazione saliente è stata che le forme *A* sembrano legarsi in maniera più efficace della forma *E*. Come risulterà evidente più avanti, importanti cambiamenti nelle quantità relative della forma *E* e delle forme *A* in differenti parti del sistema nervoso in via di sviluppo sono interpretabili alla luce di questa diversità nella capacità di legarsi.

La modulazione locale della superficie cellulare altererebbe in tre modi il legame tra cellule. La prevalenza di una particolare CAM sulla superficie di una cellula potrebbe modificarsi *(1)*, così come la sua distribuzione sulla superficie *(2)*. Oppure la molecola stessa potrebbe venire modificata chimicamente *(3)*, come capita nel passaggio dalla forma embrionale *(E)* all'adulta *(A)* della N-CAM.

Il tasso di aggregazione di molecole lipidiche artificiali che possiedono la forma *E* *(in nero)* ed *A* *(in colore)* della N-CAM è stato misurato da Stanley Hoffman. I risultati suggeriscono che esso aumenti per due forme di modulazione della superficie: la conversione da *E* ad *A* ha dato luogo a un aumento di cinque volte, mentre la prevalenza raddoppiata di una forma ha condotto a un aumento di 30 volte.

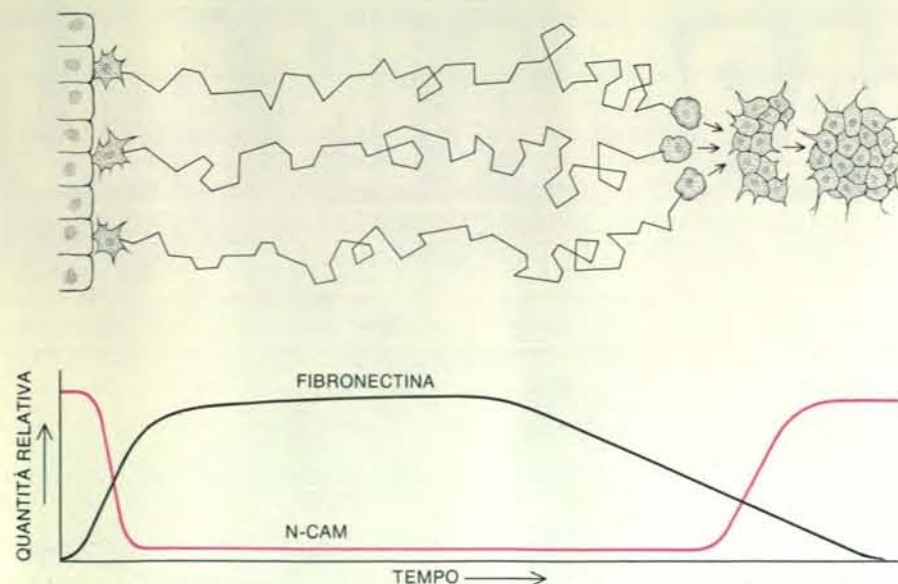
Le strutture particolareggiate delle altre due molecole per l'adesione tra cellule, che sono state identificate, sono state meno ben determinate, ma queste molecole appaiono comunque chiaramente diverse l'una dall'altra oltre che dalla N-CAM. La L-CAM ha un peso molecolare di 124 000, inferiore a quello della N-CAM, e manca dell'acido polisialico insolito. Diversamente dalla N-CAM, essa media l'adesione tra cellule solo quando sono presenti ioni calcio. Non si sa bene se il legame della L-CAM sia omofilo. Studi sulla struttura e sulla scissione di questa molecola indicano che la sua porzione proteica è anch'essa notevolmente diversa da quella della N-CAM. Benché la L-CAM sia stata isolata per la prima volta dalle cellule epatiche, essa svolge, durante lo sviluppo precoce dell'embrione, una parte fondamentale nella congiunzione con la N-CAM.

La terza molecola per l'adesione, isolata in epoca molto recente, è la Ng-CAM, che compare più tardivamente nello sviluppo e media il legame dei neuroni a cellule di sostegno presenti nel sistema nervoso, note come glia. La componente principale della Ng-CAM ha un peso molecolare di 135 000; due altre componenti, con peso molecolare maggiore e minore, vengono anch'esse isolate dagli anticorpi prodotti contro questa CAM. La Ng-CAM si trova sui neuroni, ma non sulla glia alla quale quei neuroni aderiscono. Presubilmente, essa si lega a una molecola per l'adesione, presente sulla glia, che dobbiamo però ancora identificare. In altre parole, il meccanismo di legame è probabilmente di tipo eterofilo.

L'espressione di due CAM è stata seguita da Thiery in studi di immunofluorescenza. Sezioni di tessuti adiacenti sono state trattate con anticorpi contro la N-CAM *(a sinistra)* e la L-CAM *(a destra)*, marcati con un colorante fluorescente. In una sezione trasversale, che mostra la placca neurale mentre si forma la doccia neurale, la N-CAM è presente in grandi quantità nell'ectoderma neurale alla base della doccia. La L-CAM è confinata nell'ectoderma non neurale a fianco della doccia. In uno stadio più precoce, il blastoderma si colora uniformemente per la N-CAM e la L-CAM.

Questa mappa presuntiva embrionale, che si basa su una mappa costruita in precedenza da Luc Vakaet dell'Università Statale di Anversa, in Belgio, mostra, su un disco che rappresenta la superficie del blastoderma, l'origine di cellule che produrranno particolari tessuti e organi *(in alto)*. Questa mappa è stata poi combinata con dati di immunofluorescenza per ottenere una mappa composita per le CAM *(in basso)*, in grado di mostrare la distribuzione della N-CAM *(in colore)* e della L-CAM *(in grigio)* e i tessuti nei quali né l'una né l'altra di queste sostanze sono presenti *(in bianco)*. Vi è per la N-CAM un gradiente poco profondo, indicato da due colori. Le cellule mesodermiche che formano le vie urinarie esprimono ambedue le CAM. Si noti che le regioni centrali dell'ectoderma neurale e alcune regioni mesodermiche che esprimono la N-CAM sono circondate da regioni che esprimono la L-CAM, cioè da ectoderma non neurale e da endoderma.

La N-CAM, la L-CAM e la Ng-CAM non risultano dotate di una specificità incrociata, cioè nessuna di esse si lega in maniera effettiva, da cellula a cellula, all'una o all'altra delle altre molecole. Rimane da chiarire se la N-CAM e la Ng-CAM sullo stesso neurone possano interagire tra loro.

Se il modello del corso d'acqua di montagna si applica al legame tra cellule e alla comparsa di una determinata conformazione, dovrebbero esservi meccanismi che alterano o il tasso o l'intensità di legame tra molecole per l'adesione tra cellule, presenti su cellule diverse. Nel 1976 ho proposto che la modulazione locale della superficie cellulare - cambiamenti nella quantità, nella distribuzione o nella chimica di particolari molecole della superficie cellulare - potesse provocare questi cambiamenti di legame, cambiamenti che dovrebbero modificare le interazioni tra cellule soggette ad altri processi primari come i movimenti morfogenetici. Le interazioni alterate che influiscono sui processi primari promuoverebbero a loro volta la formazione di differenti configurazioni. Abbiamo trovato, di fatto, che durante l'embriogenesi precoce nelle varie CAM hanno luogo variazioni riguardo alla prevalenza e che una importante riduzione nella quantità di acido sialico, la conversione da *E* a *A*, ha luogo nella N-CAM durante lo sviluppo perinatale (all'incirca al momento della nascita nel topo e della schiusa nel pollo).

Per dimostrare che tali cambiamenti nella modulazione potrebbero davvero influire sul legame avevamo bisogno di un saggio molecolare diretto. Stanley Hoffman e io stesso abbiamo deciso di misurare la velocità di aggregazione di varie combinazioni di vescicole lipidiche artificiali in cui erano state inserite o la forma *E* o le forme *A* della N-CAM. La nostra previsione era che le vescicole *E* si sarebbero aggregate l'una all'altra a una velocità inferiore di quella alla quale si sarebbero aggregate alle vescicole *A* e che queste si sarebbero aggregate invece tra di loro alla velocità massima. Tale previsione era basata sul fatto che l'acido sialico, nell'area intermedia, o altera la forma dell'area che si lega o direttamente respinge la molecola di N-CAM su un'altra cellula, che si trova di fronte ad esso; quanto minore è la quantità di glucide presente, tanto più elevata dovrebbe essere la velocità di aggregazione. La previsione è stata chiaramente confermata.

Ancora più sorprendente è stato l'effetto determinato dall'aumento della N-CAM di una determinata forma sulla membrana. Un aumento di due volte ha determinato un aumento superiore alle trenta volte nei tassi di legame. Sia la conversione da *E* a *A* sia i cambiamenti nella prevalenza in superficie potrebbero dunque presumibilmente portare, nel corso dello sviluppo, a notevoli variazioni nei tassi di aggregazione, una condizione necessaria per qualunque modello di formazione di una configurazione che abbia dei limiti cinetici. È importante riconoscere che ambedue questi cambia-

La migrazione delle cellule è associata a cambiamenti nell'espressione delle CAM, come è stato dimostrato in uno studio sulle cellule della cresta neurale, realizzato da Thiery. Queste cellule hanno origine alla sommità del tubo neurale e migrano diffusamente formando varie strutture tra cui gangli del sistema nervoso autonomo e sensitivi. Nella parte in alto sono schematizzati i percorsi tipici di tre di queste cellule quando si allontanano dal tubo neurale e finiscono per congiungersi con altre cellule, formando un ganglio *(a destra)*. Una curva mostra in che modo la N-CAM, che è presente all'origine sulle cellule, scompaia dalla loro superficie quando esse si muovono e ricompaia, invece, quando raggiungono la loro destinazione e interagiscono con altre cellule. L'altra curva mostra un aumento e una caduta coordinati di fibronectina, una molecola che funge da substrato e che, come Thiery aveva già mostrato, forma un tappeto sul quale si spostano le cellule.

menti nella modulazione siano essenzialmente continui e perciò definiscono un vastissimo numero di possibili stati di legame per la N-CAM. Analoghe attese si hanno per le CAM di altre specificità.

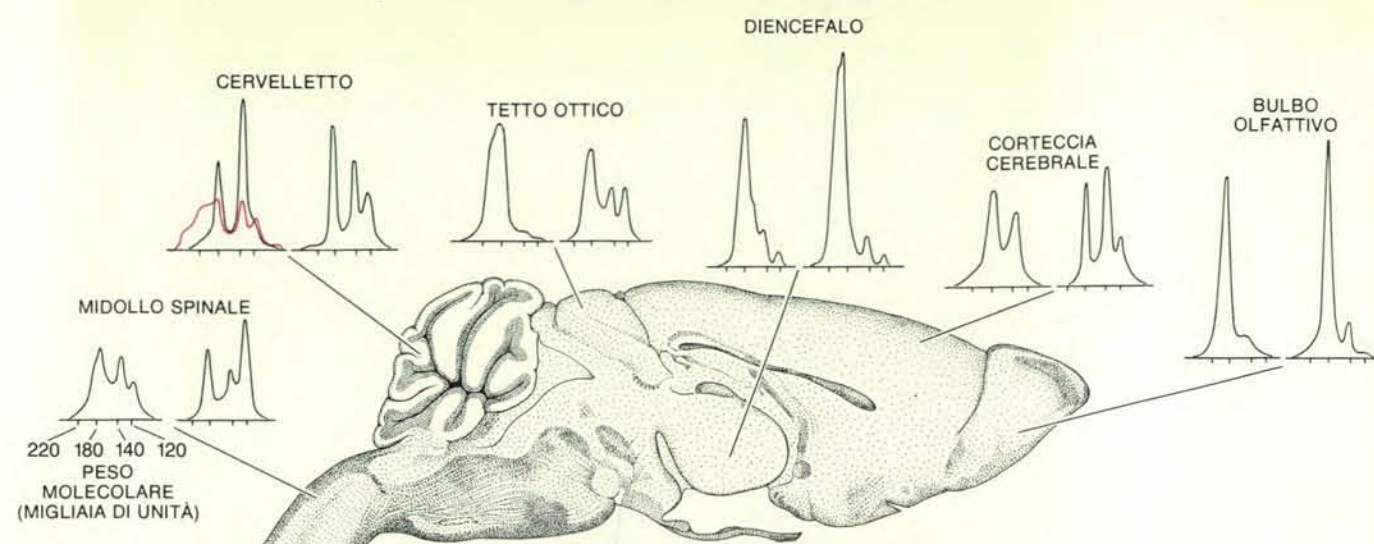
In quale modo i cambiamenti nelle CAM si correlano con alterazioni nella forma dell'embrione? Jean-Paul Thiery, che lavora oggi all'Institut d'Embryologie di Nogent-sur-Marne, in Francia, e io stesso abbiamo trovato che sia la N-CAM sia la L-CAM si possono individuare molto precocemente nell'embrione di pollo. Immediatamente prima della formazione dei foglietti embrionali, l'epiblasto e l'ipoblasto si colorano in modo più o meno uniforme con anticorpi fluorescenti contro queste CAM. In seguito, allorché si sviluppa la linea primitiva, le cellule dello strato intermedio che stanno migrando non si colorano. A questo stadio dello sviluppo si osserva tutta una serie di importanti transizioni. Le cellule che daranno origine alla placca neurale non si coloreranno più con gli anticorpi contro la L-CAM, ma si coloreranno intensamente con gli anticorpi contro la N-CAM. Un cambiamento correlato ha luogo nella regione circostante: si sviluppa un bordo tra l'ectoderma neurale e quello non neurale e le cellule non neurali si colorano soltanto con gli anticorpi contro la L-CAM, proprio come fanno le cellule dell'endoderma. La comparsa di questo tipo di segregazione e la formazione di un bordo accompagnano l'evento chiave dell'induzione neurale.

Che cosa questi e i successivi cambiamenti nella distribuzione delle CAM potrebbero significare per la definizione della configurazione? Gli embriologi costruiscono «mappe presuntive», che mostrano che cosa succederà alle cellule di ogni regione dell'embrione, indicando i tessuti e le strutture ai quali esse daranno origine in un ben definito periodo di tempo. Luc Vakaet dell'Università Statale di Anversa, in Belgio, ha costruito una mappa presuntiva embrionale particolarmente accurata, che mostra quanto accade alle cellule del blastoderma di pollo nel momento in cui la formazione degli organi è completa. Noi abbiamo rintracciato l'espressione di quelle cellule e delle loro discendenti riguardo alle CAM e abbiamo costruito una mappa presuntiva composita, combinando i nostri risultati con quelli di Vakaet *(si veda la figura in basso a pagina 92)*. Questa mappa ha numerosi aspetti interessanti: per capirli occorre rendersi conto che una distribuzione tetradimensionale (considerando il tempo e le tre dimensioni dello spazio) delle cellule e dei marcatori CAM viene rappresentata sotto forma di mappa su una superficie bidimensionale, che rappresenta la lamina costituita dalle cellule progenitrici del blastoderma. Il risultato è una mappa topologica anziché una mappa topografica, con alcuni particolari strutturali esatti che sono stati sacrificati per mettere in rilievo l'associazione e la contiguità, due relazioni che risultano particolarmente importanti per gli eventi dell'induzione embrionale.

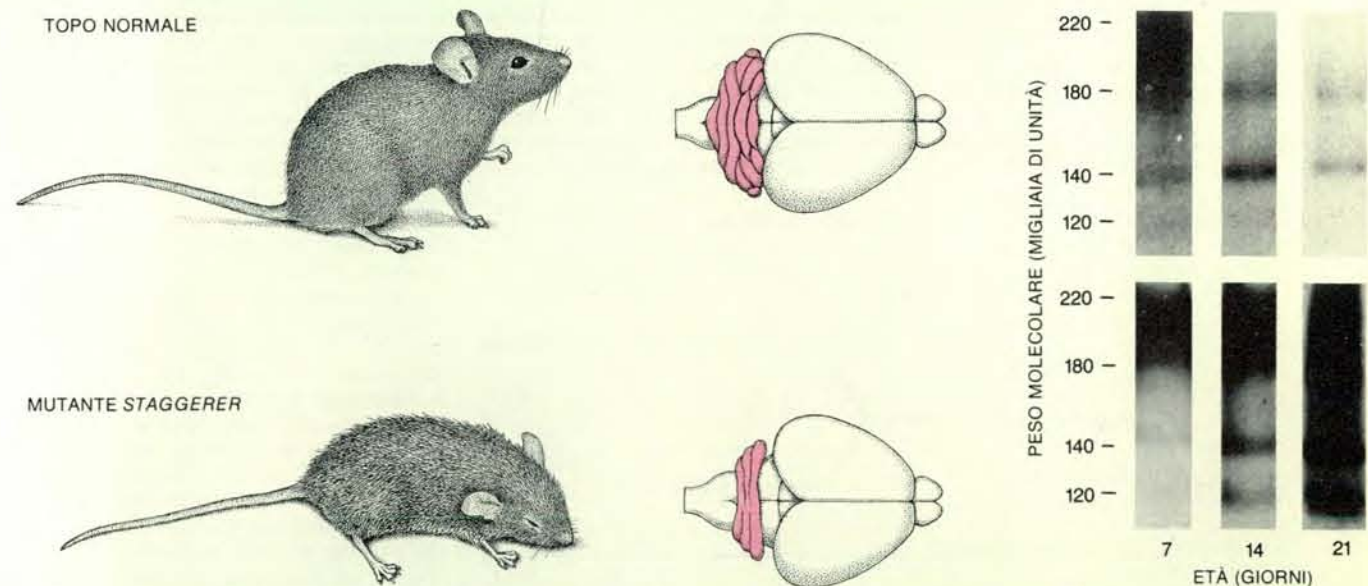
Il primo aspetto che sorprende è che le regioni N-CAM più centrali, destinate a dare origine alla placca neurale, alla notocorda, ai somiti e a certe parti del mesoderma della placca laterale sono completamente circondate da un anello contiguo e coerente di regioni che esprimeranno l'L-CAM. Queste regioni tutte insieme comprendono l'ectoderma non neurale e l'endoderma. In secondo luogo, vi è un gradiente poco profondo testa-coda di colorazione per la N-CAM. La colorazione risulta più intensa nella regione della placca neurale; nelle regioni sottostanti, come la notocorda, essa è meno intensa e l'insieme appare dinamico: all'inizio non vi è colorazione per la N-CAM, quindi la notocorda (una struttura mesodermica) si colora intensamente e, alla fine, il colore scompare. Un'analoga sequenza di comparsa e scomparsa della N-CAM si osserva nei somiti.

Una terza sequenza importante si osserva negli elementi del rene che hanno origine nel mesoderma: la L-CAM e la N-CAM compaiono e scompaiono in sequenze corrispondenti agli stadi finali dell'induzione embrionale reciproca di una massa di cellule (il mesenchima del mesonefro) da parte del canale di Wolff. In primo luogo, sul canale di Wolff compare la L-CAM; quindi, sui tubuli del mesonefro, allorché si organizzano in epiteli a partire da cellule del mesenchima, compare la N-CAM, che viene poi sostituita dalla L-CAM allorché i tubuli in un secondo tempo si dilatano. Infine, la mappa presenta regioni che non si colorano affatto con gli anticorpi contro la N-CAM o la L-CAM. Questo fatto suggerisce l'idea che perlomeno un'altra CAM a comparsa precoce, finora non ancora identificata, medi le interazioni che portano all'adesione delle cellule tra loro in strutture che derivano da queste aree durante le prime fasi dello sviluppo.

Si fanno strada due generalizzazioni. Una è che le CAM subiscono cambiamenti dinamici nell'ordine della loro comparsa, distribuzione e quantità dovunque siano in corso induzioni primarie e secondarie. L'altra è che, dovunque gli epiteli si trasformano in mesenchima, sembra che le superfici cellulari perdano le CAM.

Un'interpretazione un poco fantasiosa della mappa suggerisce in che modo possano comparire alcune configurazioni particolari. Se si ammette che sia il legame con la N-CAM sia quello con la L-CAM siano omofili, allora, dati i noti movimenti morfogenetici, si può vedere dalla mappa come può ripiegarsi la placca neurale a formare un tubo neurale chiuso a seguito di interazioni con la N-CAM e come le interazioni omotipiche tra cellule endodermiche, all'interno, possano formare l'intestino. La mappa suggerisce anche in che modo strutture come l'intestino e il faringe possano entrare in contatto con la cute, anche se hanno origine in differenti foglietti embrionali: quello che hanno in comune sono le L-CAM. Che queste congetture siano vere o no, la mappa rivela chiaramente la topologia dell'espressione delle CAM e il fatto che nessuna di esse è confinata in un unico foglietto embrionale. Questi risultati contribuiscono a spiegare parecchi determi-

Il tipo di conversione da *E* ad *A* varia in differenti parti dell'encefalo del topo. L'espressione differenziale delle forme embrionale e adulta di N-CAM è qui illustrata per cinque regioni da due curve. Ogni curva rappresenta un'analisi densitometrica di un gel elettroforetico su cui la CAM estratta da una regione è stata frazionata. La curva a sinistra si riferisce a topi di 21 giorni, quella a destra a topi di 180 giorni. La densità ottica (e così la quantità di CAM) è indicata sull'asse verticale e il peso molecolare sull'asse orizzontale; la forma embrionale *(E)* è rappresentata da un apice a più di 200 000 e le tre forme adulte *(A)* da apici a 180 000, 140 000 e 120 000. Nella maggior parte delle regioni compare pochissima forma *E* a 21 giorni e meno o niente affatto a 180 giorni, mentre la scansione temporale e l'entità della conversione sono diverse nelle varie regioni. Nel topo mutante *staggerer*, la conversione nel cervelletto è dilazionata *(curva in colore)*: a 21 giorni, nelle cellule cerebellari vi è ancora una notevole quantità della forma *E* della N-CAM e meno del normale delle forme *A*.

nanti della forma embrionale e suggeriscono che la distribuzione delle CAM e le loro sequenze di espressione abbiano un'importanza critica. Benché le CAM non specifichino soltanto la forma, è probabile che esse pongano i limiti necessari a quei processi che servono a modellare l'embrione.

Quante CAM sono necessarie per riempire una mappa presuntiva? Finora due CAM rispondono per più di due terzi della superficie di un embrione precoce. Come ho già ricordato, una terza CAM primaria, o addirittura una quarta, può essere necessaria per completare la mappa, ma in ogni caso sembra probabile che il numero richiesto per l'embriogenesi precoce non sia molto grande. Una questione correlata è quella di sapere se CAM supplementari (o differenti forme di CAM note) vengano chiamate in causa in seguito, come responsabili di raffinate interazioni tra tessuti, quali sono quelle che si osservano negli stadi più tardivi dello sviluppo encefalico. La risposta per l'encefalo sembra essere che, pur continuando la N-CAM a essere attiva, in questi stadi tardivi, nelle interazioni neurone-neurone, perlomeno una nuova molecola, la Ng-CAM, è necessaria per le interazio-

Un mutante *staggerer* viene qui confrontato con un topo normale. Un omozigote per questa mutazione (cioè un individuo che abbia ereditato la mutazione da ambedue i genitori) mostra atassia, cioè ha una postura anomala, si muove in modo irregolare e perde l'equilibrio. Questo genere di topi ha anche una costituzione più piccola e più debole della norma, non è in grado di ripulirsi e muore all'età di circa 30 giorni. Alterazioni nelle connessioni cellulari a livello del cervelletto si riflettono in una struttura grossolana del cervelletto stesso, particolarmente piccola e disordinata *(in colore)*. I gel elettroforetici *(a destra)* documentano il ritardo nella conversione dalla forma embrionale *E* alle forme adulte *A* della N-CAM nel mutante *staggerer*. Nel cervelletto del topo normale la forma *E*, ad alto peso molecolare, è stata soppiantata dalle forme *A* della N-CAM all'età di 14 giorni; nello *staggerer* la forma *E* persiste sul gel come banda diffusa anche all'età di 21 giorni.

	EMBRIONE PRECOCE	ADULTO
L-CAM	Ectoderma	Cute (strato germinativo)
	Mesoderma	Epitelio delle vie urogenitali
	Endoderma	Epiteli delle vie digerenti e respiratorie, organi linfoidi e ghiandole
N-CAM	Ectoderma	Sistema nervoso
	Mesoderma	Muscolo cardiaco; testicolo

Le CAM persistono nella vita adulta del pollo nei tessuti che derivano dai foglietti embrionali nei quali ciascuna di esse compare in stadio precoce: in tutti e tre la L-CAM, solo in due la N-CAM.

ni eterotipiche dei neuroni con la glia allorché il tessuto nervoso si sviluppa.

Mi rivolgo ora a questi eventi tardivi nella formazione del tessuto nervoso e, in maniera specifica, a uno studio sulla formazione di una configurazione nell'encefalo di topo. Questo studio ha fornito il sostegno finora più valido all'ipotesi della modulazione della superficie e al ruolo delle CAM nella formazione di un organo in un animale integro. Abbiamo pensato che, proprio come la prova di cambiamenti nella prevalenza in superficie delle N-CAM si vede riflessa nella mappa presuntiva precoce, così in seguito, in seno a un tessuto, si dovrebbe vedere l'espressione di altri tipi di modulazione. Ne sarebbero un esempio le variazioni nell'entità della conversione *E-A* in regioni neurali strutturalmente diverse. Una simile modulazione chimica, benché sia locale per un organo e sia epigenetica, dovrebbe essere nondimeno sotto controllo genetico.

Chen-Ming Chuong e io stesso abbiamo trovato che lo schema di conversione *E-A* varia di fatto da una regione istologica del cervello di topo a un'altra. Quantità relative delle due forme di N-CAM, grossolanamente diverse, sono presenti in differenti regioni, in cui la conversione da *E* a *A* inizia in momenti diversi oppure procede con una differente velocità o in diversa misura. La conversione, che è un evento epigenetico, potrebbe derivare, in una particolare regione, dalla scissione enzimatica di una parte dell'acido sialico delle CAM presenti sulla superficie cellulare. In alternativa, essa potrebbe derivare da un ricambio di CAM alla superficie cellulare, con la forma *E* che viene sostituita da una forma *A*, alla quale, da parte di un enzima intracellulare, sono state legate quantità più piccole di acido sialico. Qualunque sia il meccanismo di questa modulazione chimica, la presenza di differenti quantità delle forme *E* e *A* modificherebbe presumibilmente, in queste regioni, l'efficienza del legame tra le cellule in via di differenziamento e lo farebbe in modi diversi durante l'istogenesi, alterando la struttura in una maniera dinamica a mano a mano che procedono sia il differenziamento cellulare sia il movimento cellulare.

Anche se la conversione da *E* a *A* è epigenetica e dipende dall'attività enzimatica, gli enzimi, come altre proteine, sono chiaramente sotto controllo genetico. Per indagare sulla genetica della regolazione e della modulazione in superficie delle CAM abbiamo lavorato su un gruppo di topi mutanti, scoperto da Richard L. Sidman della Harvard Medical School. Questi mutanti presentano tutti alterazioni nello sviluppo del cervelletto; un sintomo importante ne è quel disturbo motorio e dell'equilibrio, che è noto come «atassia» e che comincia ad apparire nel periodo perinatale. Dei tre mutanti ampiamente studiati, lo *staggerer* (che barcolla) presenta disturbi a carico delle connessioni dei neuroni tra loro; il *reeler* (che gira su sé stesso) e lo *weaver* (che si muove a zig zag) hanno, invece, disturbi a carico delle connessioni nervose, che interessano anche le cellule gliali, le quali - come ha potuto dimostrare Pasko Rakic che lavora alla Yale University School of Medicine - svolgono una funzione nella migrazione delle fibre nervose durante lo sviluppo normale.

INDUZIONE NEURALE
INDUZIONI SECONDARIE
CAMBIAMENTI PERINATALI
N-CAM
L-CAM
N-CAM
L-CAM
N-CAM (*E*) → (*A*) →
Ng-CAM →
TEMPO →

La sequenza temporale dell'espressione delle CAM viene regolata sia geneticamente sia epigeneticamente. La N-CAM e la L-CAM divergono per quanto riguarda la loro distribuzione nelle cellule. La loro prevalenza è modulata (talvolta fino al punto che scompaiono), in particolare in corrispondenza dei siti dove si svolge l'induzione embrionale. Durante la formazione in dettaglio del tessuto nervoso, una molecola di un gruppo secondario, la Ng-CAM, fa la sua comparsa. Nel periodo perinatale vi sono ulteriori modulazioni epigenetiche come la conversione da *E* ad *A*.

I sintomi dello *staggerer* e difetti grossolani si notano solo negli animali omozigoti (quelli in cui i geni ereditati da entrambi i genitori portano la mutazione cruciale). In questi topi le sinapsi tra le fibre parallele della corteccia del cervelletto e i dendriti fittamente ramificati delle cellule di Purkinje (che sono anch'esse difettose nei topi *staggerer*) sono imperfette e, chiaramente, come conseguenza di questa non giusta connessione, le cellule granulari del cervelletto, che danno origine alle fibre parallele, muoiono in grandi quantità. Il risultato di tali anomalie è un animale affetto da atassia, con un cervelletto piccolo e alterato.

Abbiamo supposto che la modulazione della N-CAM fosse probabilmente difettosa nello *staggerer* e, di fatto, abbiamo trovato che le cose stavano proprio in questi termini: la conversione da *E* a *A* è fortemente ritardata nel cervelletto degli omozigoti, anche se la proteina N-CAM appare in sé normale. Al contrario, i mutanti *reeler* e *weaver* seguono un corso normale nella conversione da *E* a *A*.

Questi risultati non offrono ancora una spiegazione della causa dell'anomalia nel mutante *staggerer*, ma per contro suggeriscono che una conseguenza del difetto genetico in questione sia una insufficiente sintesi o attività degli enzimi responsabili della conversione. Negli animali normali quest'ultima può far aumentare il livello di adesione e pertanto portare a termine certe migrazioni cellulari. Un difetto nella conversione porterebbe dunque a un coordinamento alterato di vari aspetti dello sviluppo, della migrazione e della formazione delle sinapsi, osservati nei processi nervosi.

Come mostrano i mutanti *reeler* e *weaver*, tuttavia, non tutti i difetti di connessione nell'encefalo (per esempio quelli che interessano la glia) devono per forza avere origine soltanto da un'insufficiente adesione tra neurone e neurone. Di fatto, c'è da aspettarsi che i neuroni abbiano in superficie molecole diverse dalla N-CAM per mediare il loro legame con la glia, la quale ha ruoli chiave di guida nell'istogenesi precoce sia della corteccia cerebellare sia di quella cerebrale. Come ho ricordato, siamo stati in grado di identificare e di isolare questa molecola: essa è la Ng-CAM. Nell'embrione di pollo essa compare sui neuroni centrali all'incirca al terzo giorno e mezzo, appena prima della definitiva comparsa della glia nel sistema nervoso centrale. Non verrebbe rappresentata con la N-CAM e con la L-CAM in una mappa presuntiva molto precoce delle CAM, costruita quando l'embrione ha tre giorni; in una mappa più tardiva comparirebbe solo nella regione del sistema nervoso. È pertanto una CAM secondaria, limitata a certi derivati di un solo foglietto embrionale, l'ectoderma.

Sarà istruttivo esaminare la Ng-CAM nei topi *reeler* e *weaver* perché sembra che o la Ng-CAM o una CAM complementare sulla glia possa mediare certi tipi di migrazione neuronale sulla glia stessa ed essere così interessata nei suddetti disordini. Ci aspettiamo di trovare una nuova CAM sulla glia non solo perché manca la Ng-CAM, ma anche perché si può pensare che l'interazione neurone-glia sia mediata da differenti CAM, cioè da un legame eterofilo. Se, invece, l'interazione neurone-glia fosse mediata da un meccanismo omofilo, vi sarebbe una confusione completa nel distinguere tra interazioni simultanee neurone-neurone e neurone-glia. Malgrado la potenza dei meccanismi di modulazione, è dunque chiaro che certe circostanze nell'istogenesi richiedono la presenza di CAM con specificità diversa. Proprio la combinazione tra specificità e modulazione contribuisce al delinearsi del cambiamento morfogenetico.

Quanto microscopico può essere il modellamento morfogenetico mediato dalle CAM? Cioè quale grado di precisione esso è in grado di realizzare all'interno di un particolare tessuto? L'encefalo è il banco di prova più squisito per i tentativi di dare una risposta a questo interrogativo. Sarà particolarmente affascinante, per esempio, ricostruire il ruolo delle CAM nel formare precisi prolungamenti della fibra nervosa quali sono quelli che dalla retina si estendono, nel cervello, verso il tetto ottico. Gli studi preliminari sono promettenti. Con Scott E. Fraser dell'Università della California, presso l'Irvine California College of Medicine, abbiamo mostrato che la mappa retino-tettale dell'anuro *Xenopus laevis* può essere fortemente distorta nell'animale vivo introducendo nel tetto ottico anticorpi anti-N-CAM.

Mi si consenta di riassumere i dati che abbiamo ottenuto prima di proporre un'ipotesi che correli la regolazione delle CAM all'espressione dei geni e allo sviluppo della forma animale. La grande maggioranza delle prove accumulate nei nostri studi depone a favore dell'idea che la modulazione locale sulla superficie cellulare sia un fattore che offre un contributo importante alla morfogenesi. Parecchie CAM sono state identificate. Esse differiscono come struttura, specificità e modo di legarsi. La meglio studiata, la N-CAM, ha un meccanismo di legame omofilo; né essa né l'altra proteina neuronale, la Ng-CAM, richiedono ioni di calcio per legarsi. Al contrario la L-CAM, che ha una specificità completamente diversa, dipende dal calcio; non è stato invece mostrato ancora in modo definito se il suo meccanismo di legame sia omofi-

Secondo l'ipotesi dell'autore, le CAM regolano il movimento morfogenetico. L'espressione di queste molecole è controllata da un gruppo di geni e viene modulata epigeneticamente per promuovere cambiamenti dinamici nell'adesione tra le varie cellule. Quest'ultima influisce sui movimenti morfogenetici che provocano l'induzione di strutture di organi in sedi particolari. Una serie separata di geni regola il differenziamento cellulare, il quale interagisce all'interno di queste strutture per modellare, nel corso dello sviluppo embrionale, le configurazioni tissutali.

lo. Queste tre CAM sono le uniche isolate fino a oggi come molecole pure. La questione del numero totale di differenti CAM rimane aperta, ma sembra improbabile che venga richiesto un repertorio molto vasto di CAM (più di 100) con specificità diversa, in particolare visto che i meccanismi di modulazione e il differenziamento cellulare possono alterare profondamente gli effetti selettivi di queste molecole sulla configurazione che viene delineandosi nel corso dello sviluppo.

Il quadro dinamico dell'espressione delle CAM che emerge sembra compatibile con il modello del corso d'acqua di montagna. Durante lo sviluppo vi è una sequenza di espressioni sia reversibili sia irreversibili dei geni delle CAM, coordinate con alterazioni epigenetiche nell'espressione superficiale di differenti CAM. Questi eventi si svolgono in punti specifici e in una particolare sequenza nelle discendenti di particolari cellule. Benché non siano noti i segnali causali di queste espressioni differenziali, è ragionevole supporre che, nel corso dell'evoluzione, la selezione abbia favorito gli individui nei quali le CAM si esprimevano in sequenze che davano una forma funzionale. Queste sequenze dovrebbero essere adatte per ogni gruppo di cellule che faccia parte della serie coordinata di movimenti morfogenetici che conducono all'induzione embrionale; ogni difetto di induzione, derivante da un imperfetto coordinamento tra l'espressione delle CAM e i cambiamenti nei movimenti sarebbe fatale. La correlazione tra la successione dalla L-CAM alla N-CAM e il verificarsi di varie induzioni primarie, secondarie e reciproche, che è stata osservata, non può essere accidentale.

Non si sa se gli eventi di adesione tra cellula e cellula esercitino un meccanismo di retroazione per influire in modo diretto sui processi primari del differenziamento cellulare o della divisione cellulare. Sembra probabile che essi abbiano perlomeno effetti indiretti che modificano il contributo proporzionale di ogni processo primario alla morfogenesi. Per esempio, un evento modellatore fondamentale nell'embriogenesi, la transizione dall'epitelio al mesenchima, può essere spiegato con la modulazione della prevalenza delle CAM, come nel caso delle cellule della cresta neurale che migrano, studiate da Thiery. In accordo con l'idea fondamentale che l'espressione differenziale precoce dei geni delle CAM sia necessaria per la morfogenesi, esistono chiaramente dei segnali per far esprimere i geni per le CAM secondarie, come la Ng-CAM, in momenti critici: immediatamente prima che siano necessari per le interazioni eterotipiche nell'istogenesi. È degno di nota il fatto che ogni CAM persista nella vita adulta solo nei tessuti derivati dal foglietto embrionale nel quale, in origine, è stata espressa. Non sorprenderebbe di trovare che le alterazioni nelle CAM sono associate con una vasta gamma di processi patologici, tra cui la metastasi di cellule neoplastiche. Robert Brackenbury, Michael E. Greenberg e io stesso abbiamo già mostrato che la N-CAM scompare dalle cellule nervose trasformate dai virus cancerogeni.

Un'ipotesi minima sulla base molecolare della morfogenesi deve prendere in considerazione tre processi primari: l'adesione tra le cellule, il movimento cellulare e il differenziamento cellulare. (Sufficienti prove sono state raccolte in differenti animali per mostrare che la divisione cellulare e la morte cellulare hanno funzioni importanti in particolari tipi di morfogenesi in determinate specie, ma è chiaro che la forma locale può talvolta essere definita senza di esse.) L'espressione precoce delle CAM primarie deve essere in genere indipendente dal differenziamento cellulare, perché una data CAM primaria compare in regioni indotte che solo in seguito formeranno tipi cellulari differenti, in organi nettamente diversi. D'altra parte, l'espressione delle CAM deve avere un ruolo diretto nel controllo del movimento morfogenetico che deriva dall'interazione tra motilità cellulare, tensione nelle lamine di tessuto e adesione tra le cellule.

Nell'ipotesi del regolatore ho suggerito che i movimenti morfogenetici, che riuniscono insieme cellule e tessuti con differenti storie nelle induzioni embrionali e determinano così le fasi successive del processo morfogenetico, siano regolati dalle sequenze di espressione delle CAM. La sintesi delle CAM sembrerebbe regolata da uno speciale gruppo di meccanismi, distinti da quelli che in seguito regoleranno il differenziamento cellulare. Ancora non sappiamo che cosa inneschi i geni regolatori delle CAM, ma, quando essi sono innescati, la serie di modulazioni delle CAM che ne consegue limita in modo marcato il movimento delle cellule e altri processi primari. Avendo lo sviluppo un decorso storico, anche le medesime sequenze di espressione dei geni per la L-CAM e per la N-CAM possono probabilmente dare origine a nuove strutture in differenti contesti di differenziamenti cellulari successivi.

Se l'ipotesi del regolatore risulta corretta, sarebbe possibile spiegare perché differenti specie di vertebrati possono avere differenti tipi di gastrulazione, movimenti morfogenetici e particolari delle loro mappe presuntive embrionali diversi e, malgrado questo, avere un piano corporeo fondamentale comune. La ragione che viene proposta è che la selezione naturale nel corso dell'evoluzione permette la sopravvivenza solo degli individui nei quali varie combinazioni di movimenti morfogenetici e di espressioni dei geni per le CAM conducono a opportune sequenze induttive funzionali.

Nuove CAM e ulteriori meccanismi di modulazione verranno senza dubbio scoperti, arricchendo notevolmente e ampliando questa visione preliminare del rapporto tra adesione tra cellule e morfogenesi. Si hanno già sufficienti nozioni per poter dire che le molecole che favoriscono l'adesione tra cellule e hanno specificità diverse, comparendo in tempi diversi nel corso dello sviluppo come risultato di un'espressione genica differenziale e di una modulazione superficiale, possono agire alla stessa stregua di limitazioni cinetiche su altri processi primari dello sviluppo, favorendo in questo modo l'affermarsi della forma animale.

La nidificazione dei dinosauri

La scoperta di grandi quantità di uova e di giovani scheletri in due località del Montana, negli Stati Uniti, ha favorito una nuova interpretazione delle relazioni sociali di questi grandi rettili estinti

di John R. Horner

Fra tutti gli animali fossili, i dinosauri sono indubbiamente i più noti. In quasi tutti i continenti sono state trovate migliaia di loro scheletri completi o parziali, distribuiti in strati che rappresentano approssimativamente 140 milioni di anni della storia della Terra; è stato quindi possibile farsi un'idea precisa del loro aspetto. Non si può dire altrettanto dei loro modi di vita e dei loro comportamenti. Tuttavia, nel corso degli ultimi cinque anni, in due siti del Montana, negli Stati Uniti, sono stati trovati scheletri e uova, associati in modo tale da fornire indizi sul comportamento sociale dei tre tipi di dinosauri ritrovati.

L'ampio scenario di quest'esposizione è il continente nordamericano di 80 milioni di anni fa, che a quel tempo era diviso in due parti da un mare poco profondo, il mare interno occidentale del Cretaceo. Le Montagne Rocciose in corso di formazione e qualche sistema vulcanico isolato costituivano i rilievi dell'«America occidentale» e gli Appalachi quelli dell'«America orientale».

Dai margini orientali delle Montagne Rocciose fino alle rive occidentali del mare interno si estendeva un'ampia pianura costiera in cui si depositavano i sedimenti erosi dalle montagne. Nella regione che attualmente corrisponde al Montana e alla parte meridionale dello stato canadese di Alberta l'ampiezza di questa pianura raggiungeva i 400 chilometri, anche se questa distanza oscillava a seconda delle variazioni del livello del mare. Lungo la costa e in un determinato settore del retroterra erano frequenti gli ampi fiumi, i delta fluviali, le paludi e gli acquitrini, con una densa vegetazione probabilmente assai simile a quella esistente oggi sulla costa meridionale della Louisiana. Nel clima subtropicale umido crescevano cipressi calvi *(Taxodium distichum)*, sequoie e numerose latifoglie.

Questo ambiente era abitato da pesci, anfibi, testuggini acquatiche, coccodrilli e piccoli mammiferi primitivi. Erano presenti anche dinosauri, rappresentati principalmente dagli adrosauri (dinosauri con il muso a becco d'anatra), dai ceratopsidi (dinosauri muniti di corna) e da numerosi piccoli e grossi carnosauri. Essendo stato accertato che essi vivevano in questa zona, è sorprendente che i sedimenti contengano solo rari e frammentari resti di uova e di giovani esemplari di dinosauri, che - a quanto sembra - sarebbero stati frantumati altrove e successivamente trasportati qui.

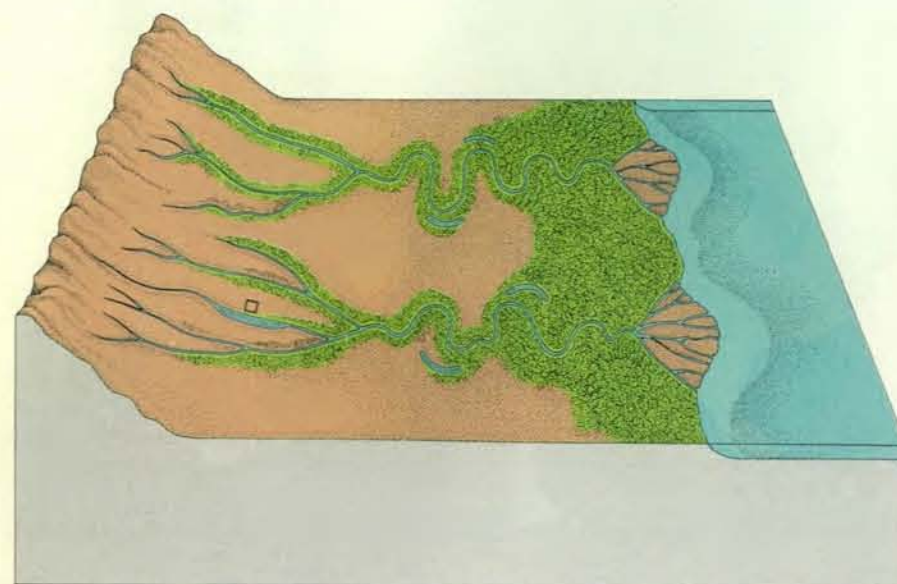
La scarsità di questo tipo di fossili in aree che chiaramente erano state frequentate dagli adulti era così paradossale che molti paleontologi hanno avanzato varie ipotesi per darne una spiegazione. Secondo alcuni, i dinosauri nidificavano in zone esposte più all'erosione che alla sedimentazione; secondo altri in territori molto discosti dalla pianura costiera, il che significa pressappoco la stessa cosa. Altri ancora attribuivano quest'anomalia alla possibilità che i dinosauri avessero un periodo di vita talmente lungo da non richiedere un gran numero di giovani individui per mantenere una determinata consistenza della popolazione adulta.

Il problema si sarebbe potuto risolvere da tempo se, all'inizio del secolo, Charles W. Gilmore, un paleontologo dello Smithsonian Institute, si fosse reso pienamente conto del significato di una delle proprie scoperte. Nel Montana occidentale egli aveva raccolto molti esemplari di dinosauri da strati della fine del Cretaceo, noti con il nome di Two Medicine Formation. Nei suoi appunti sul campo egli aveva annotato l'abbondanza di frammenti di uova. Inoltre, molti fossili di dinosauri, da lui raccolti e descritti come rappresentativi di nuove specie, sono stati successivamente riconosciuti come appartenenti a giovani esemplari di specie già note da sedimenti situati più a oriente.

I sedimenti della Two Medicine Formation differiscono da quelli contenenti i resti di dinosauri adulti. Essi vennero depositati presso le giovani Montagne Rocciose negli alvei intercomunicanti dei corsi d'acqua e nelle singole zone di straripamento; grandi quantità vennero depositate anche nei laghi. Lo studio delle rocce conferma che di norma, in determinati periodi dell'anno, il clima era asciutto. L'aspetto sorprendente di questi numerosi resti animali è l'abbondanza di uova intatte di dinosauri, di loro parti e di scheletri di giovani dinosauri. I frammenti di gusci d'uova sono i fossili più frequenti nei sedimenti e quasi l'80 per cento dei resti scheletrici appartiene a esemplari giovani di dinosauri.

È necessario tener presente che la Two Medicine Formation ha uno spessore di 650 metri, con province sedimentarie differenziate che rappresentano parecchi ambienti. Il sito su cui ho concentrato il mio lavoro corrisponde a un unico ambiente, mentre la zona di raccolta di Gilmore ne riguardava un altro. Ambedue avevano più caratteri in comune tra loro che con gli ambienti situati a oriente, ma tuttavia essi erano ancora notevolmente diversi.

Il sito dove ho operato si chiama Willow Creek Anticline. Esso si estende per meno di due chilometri e ha una sezione verticale esposta di circa 150 metri. È stato scoperto nel 1978 da Marion Brandvold, una paleontologa dilettante di Bynum, nel Montana, che aveva trovato un oggetto, poi riconosciuto come un nido contenente giovani adrosauri. Durante gli ultimi cinque anni, le raccolte condotte dalla Princeton University e dalla Montana State University, finanziate in parte dalla National Science Foundation, hanno portato alla luce i resti di circa 300 uova di dinosauro, sia intatte sia parziali, attribuite ad almeno tre specie diverse; i resti di oltre 60 scheletri completi o parziali di dinosauri; numerosi resti di lucertole e di testuggini, un grosso pterosauro («lucertola volante») e abbondanti resti di invertebrati terrestri. In questo sito, a differenza di altri in cui sono stati trovati resti di dinosauri, le uova si trovano per la

L'illustrazione di Doug Henderson, che l'ha realizzata in stretta collaborazione con l'autore, è un tentativo di rappresentazione dello scenario che si presentava ai piccoli adrosauri appena nati, dall'interno del loro nido. Secondo l'autore, gli elementi di prova raccolti da lui e dai suoi colleghi indicano che i giovani dinosauri rimanevano nel nido fino a quando avevano raggiunto l'età opportuna per avventurarsi all'aperto e per andare alla ricerca del cibo con gli adulti. Nel frattempo, la dieta principale dei piccoli che non erano in grado di nutrirsi da soli era probabilmente costituita da semi, bacche, frammenti di arbusti e altre forme di vegetazione che uno o entrambi i genitori portavano nel nido. I piccoli misuravano circa 30 centimetri di lunghezza al momento della nascita; gli adulti invece raggiungevano all'incirca i sette metri.

Durante il Cretaceo, circa 80 milioni di anni fa, un mare poco profondo divideva in due parti il continente nordamericano. L'ampia pianura costiera, che si estendeva dai margini orientali delle Montagne Rocciose, appena sollevatesi, fino alla sponda occidentale di questo mare interno, era abitata da molte specie di animali primitivi, fra cui i dinosauri, i cui resti fossili sono stati recentemente trovati presso la Willow Creek Anticline del Montana occidentale, nell'area delimitata dal rettangolo che appare sulla cartina in alto. In basso, la veduta prospettica del paesaggio dell'America occidentale durante il Cretaceo localizza il sito ancora con un rettangolo.

maggior parte nell'esatta posizione in cui furono deposte e gran parte degli scheletri, da quanto è stato possibile determinare, si trova nella postura e nel luogo occupato dall'animale al momento della sua morte. Fatta eccezione per un solo punto del sito, non sembra che questi resti siano stati smembrati o dispersi dall'azione dei corsi d'acqua.

Uno dei tre tipi di uova di dinosauro rinvenuti nei sedimenti della Willow Creek Anticline è stato attribuito agli adrosauri, perché i frammenti sono stati trovati in nidi contenenti fossili di giovani esemplari di questi animali. Secondo la ricostruzione effettuata, ogni uovo era lungo circa 20 centimetri e aveva una forma ellissoidale asimmetrica. La superficie era percorsa da rilievi, o creste, seghettati con le zone più basse orientate nel senso della lunghezza. La maggior parte dei frammenti aveva una giacitura circolare, indicazione questa che le uova venivano disposte in un nido di forma circolare.

Il secondo tipo di uovo misura circa 15 centimetri, ha una forma più allungata e striature in senso longitudinale appena accennate. Queste uova, che appartengono a un dinosauro bipede primitivo simile a *Hypsilophodon*, si trovano in covate circolari che possono raggiungere il numero di 24 per nido. Le covate hanno un diametro di circa un metro e le uova sono immerse nel sedimento, in posizione verticale o obliqua, con la parte più stretta verso il basso; inoltre sono distribuite in modo tale, che non entrano reciprocamente in contatto, salvo nei casi in cui è evidente che sono state spostate.

Il terzo tipo di uovo è di origine sconosciuta: è lungo circa 10 centimetri e ha una forma ellissoidale più regolare. La sua superficie è nodulare, senza un evidente orientamento preferenziale delle depressioni. Le uova di questo tipo sono disposte orizzontalmente in file lineari accoppiate e sembrano essere state completamente coperte dai sedimenti, a differenza di quelle rinvenute nei nidi circolari, che probabilmente penetravano nel sedimento solo in parte.

Richard Mellon, che a quell'epoca era laureando a Princeton, ha esaminato le varie strutture superficiali delle uova in relazione alla disposizione che queste avevano nelle covate, ed è arrivato alla conclusione che esse sarebbero adattamenti finalizzati alla dispersione dei gas. Il metabolismo di un embrione libera anidride carbonica attraverso le porosità del guscio dell'uovo, ma quando questo è parzialmente o completamente coperto dal sedimento deve essere elaborato un meccanismo qualsiasi perché il gas possa fuoriuscire e disperdersi in superficie. Le creste e i noduli potrebbero essere serviti a questo scopo, mantenendo il sedimento discosto dai pori.

Per quale ragione allora alcune uova presentano un orientamento delle creste o delle striature e altre una tessitura nodulare priva di orientamento preciso? Mellon ha fatto osservare che le uova ellissoidali asimmetriche sono sempre state rinvenute in posizione più o meno verticale, in nidi circolari, e che si tratta sempre di uova con creste o striature. Per quelle di minori dimensioni, del tipo attribuito agli ipsilofodonti, erano probabilmente sufficienti striature poco accentuate, in quanto l'area superficiale in contatto con il sedimento era minima. Le uova di adrosauro avevano una maggior superficie esterna e pertanto può darsi che, per poter liberare efficacemente il gas, esse dovessero avere rilievi più pronunciati. Le uova che venivano coperte totalmente dal sedimento avrebbero richiesto strutture sulla superficie esterna, indipendentemente dalle loro dimensioni. Le grosse uova trovate in Francia, e attribuite al brontosauro *Hypselosaurus*, hanno una superficie nodulare semplice. (Sembra anche che venissero disposte in file rettilinee.)

Tutte le covate conosciute e attribuite ai dinosauri hanno disposizione circolare o lineare. Per quale ragione? La risposta potrebbe essere ricercata nella struttura sociale della particolare specie in esame.

Le uova disposte in covate circolari sono concentrate in uno spazio ristretto, mentre quelle disposte linearmente sono sparse su un'area più grande. Se - come dimostrano i reperti - le covate circolari erano coperte solo parzialmente dal sedimento, i genitori (o uno solo di essi) dovevano sorvegliarle per poterle incubare e proteggere dai predatori. Questa funzione incideva sicuramente in misura sensibile sul bilancio energetico del genitore, ma il tasso di sopravvivenza delle uova era probabilmente aumentato.

Per favorire l'incubazione delle covate a disposizione circolare, la parte esposta delle uova veniva probabilmente coperta da frammenti di origine vegetale come avviene oggi in molte specie di coccodrilli moderni e di uccelli che nidificano al suolo. Questo materiale in decomposizione è soggetto a fermentazione e produce una certa quantità di calore, che può essere regolata da opportune aggiunte o rimozioni di materiale. Le uova deposte linearmente venivano probabilmente incubate con il calore prodotto dal terreno riscaldato dal sole e il successo del loro

Il sito di nidificazione degli adrosauri nella Willow Creek Anticline viene scavato e analizzato dai componenti del gruppo di ricerca dell'autore allo scopo di recuperare ogni possibile frammento di uovo o parti di scheletro giacenti alla superficie o vicino ad essa. I cerchi in colore sul terreno indicano la localizzazione dei nidi. La Willow Creek Anticline, il sito dove ha operato l'autore, si estende per meno di due chilometri e ha una sezione verticale esposta di 150 metri. La sua scoperta risale al 1978 ed è dovuta a una paleontologa dilettante di Bynum nel Montana, Marion Brandvold. Nei successivi cinque anni di ricerche, finanziate in parte dalla National Science Foundation e condotte da un gruppo della Princeton University e della Montana State University, sono venute alla luce nel sito circa 300 uova, tra integre e in frammenti, e i resti di oltre 60 scheletri di dinosauri, alcuni dei quali integri.

Nei siti del Montana sono stati trovati tre tipi di uova di dinosauro. La grossa forma ovoidale a sinistra è una ricostruzione di un uovo di adrosauro, effettuata in base a frammenti di guscio recuperati dai nidi assieme a resti scheletrici di giovani esemplari. L'uovo è lungo circa 20 centimetri e presenta sulla sua superficie delle creste sporgenti, orientate longitudinalmente. L'uovo alquanto più piccolo e allungato al centro sembra quello di un dinosauro bipede primitivo, classificato come ipsilofodonte; la sua superficie è percorsa da striature longitudinali meno sporgenti. L'uovo a destra, di origine ignota, di forma ellissoidale più regolare, ha una superficie nodulare ed è lungo circa 10 centimetri. I due primi tipi di uova vengono tipicamente trovati in posizione verticale e solo parzialmente immersi nel sedimento; le uova del terzo tipo si presentano di solito in posizione orizzontale e sono disposte in file lineari appaiate, coperte dal sedimento.

sviluppo era strettamente dipendente dalle condizioni climatiche favorevoli. È interessante notare che molte uova disposte linearmente risultano non sviluppate, al contrario di quanto si verifica nella maggioranza di quelle che appartengono a covate con disposizione circolare.

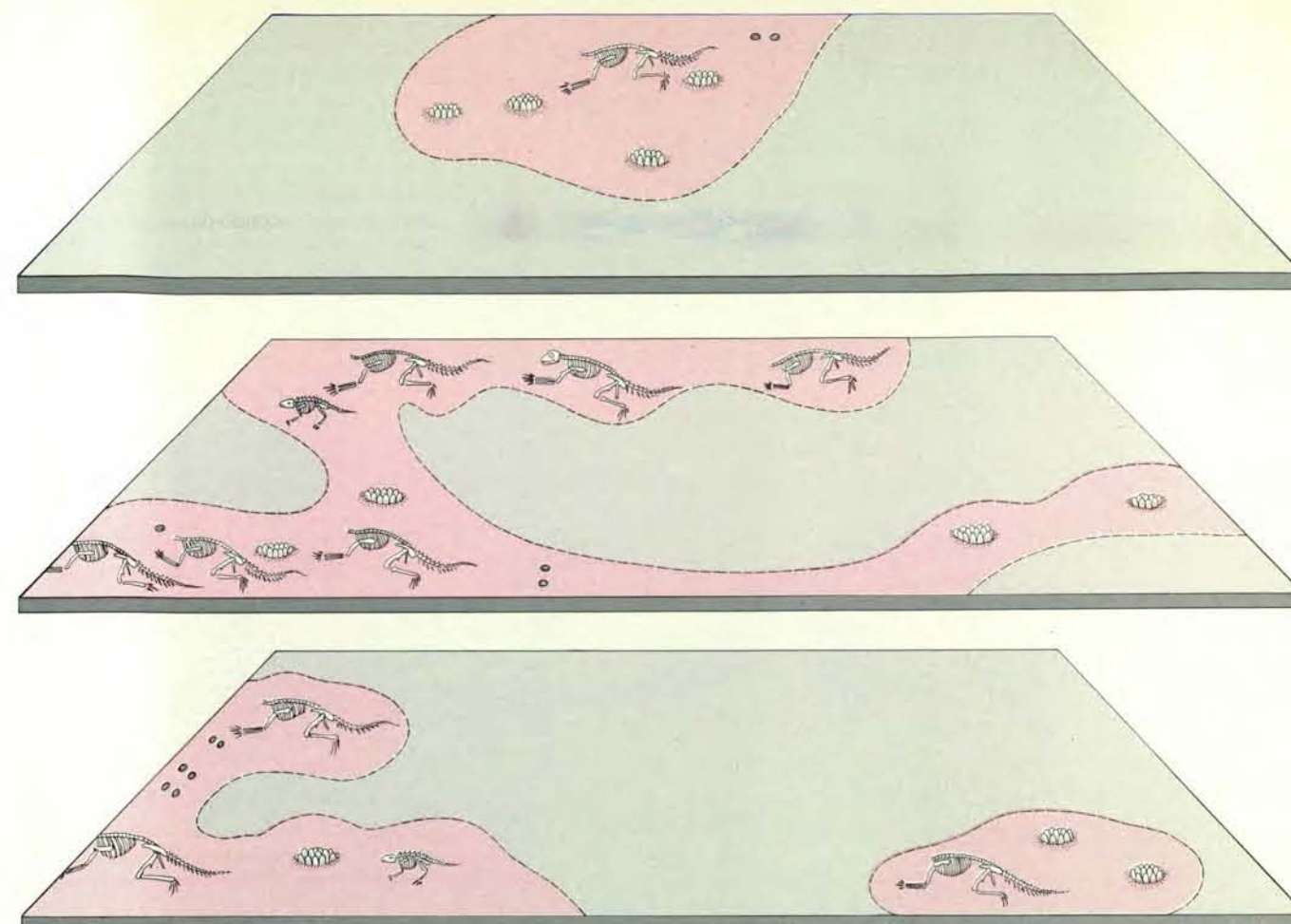
Va anche rilevato che i nidi di una determinata specie nella zona della Willow Creek Anticline sembravano riuniti in gruppi o colonie. Per esempio, le covate degli ipsilofodonti si trovano esclusivamente su due piccole alture (scherzosamente denominate «Egg Mountain» e «Egg Island»). Sulla prima tutti i nidi sono concentrati entro un diametro di circa 70 metri, contenuti in tre diversi orizzonti sedimentari e distribuiti in una sezione verticale di tre metri. I nidi di un singolo strato sono separati l'uno dall'altro da un intervallo di circa due metri; i dinosauri adulti di questa specie erano lunghi circa due metri e quindi i loro nidi dovevano essere piuttosto serrati. Il fatto che essi si trovino in strati diversi fa ritenere che si trattasse di una zona di nidificazione utilizzata per un certo arco di tempo.

I resti scheletrici trovati in loco appartengono a un gruppo composto da 20 a 25 ipsilofodonti di diverse età e i resti di individui appena nati non sono stati trovati nei nidi, ma sparsi fra essi. Il fondo delle uova contenute nei nidi è intatto e non schiacciato ed è quindi assai improbabile che i piccoli venissero allevati all'interno di tali strutture, altrimenti i gusci sarebbero stati danneggiati. Per quale ragione, allora, nella zona di nidificazione sono stati trovati individui di dimensioni diverse? Se i piccoli avessero avuto l'abitudine di abbandonare la zona appena dopo la schiusa, poteva darsi che una coppia morisse casualmente prima di allontanarsi, ma è poco verosimile che 20 o 25 decessi si verificassero contemporaneamente. Può darsi che i piccoli siano rimasti nella zona a gruppi, come avviene per molti uccelli che nidificano sul terreno. I sedimenti fanno ritenere che, quando la zona venne utilizzata per la nidificazione, essa doveva trovarsi o su una penisola protesa in un grande lago o su un'isola. Forse il cibo per i piccoli dinosauri era direttamente disponibile nel lago; perciò essi non avevano alcuna necessità di ampliare il loro raggio di ricerca degli alimenti.

Si stanno effettuando scavi, strato per strato, nella Egg Mountain, per poter determinare, per singoli periodi di tempo, le comunità animali che la hanno abitata. Nei sedimenti adiacenti alla maggior parte dei nidi di ipsilofodonti si trovano agglomerati di piccole strutture oblunghe, simili a uova, assomiglianti alle pupe dei coleotteri necrofagi moderni. Se di questo effettivamente si tratta, i coleotteri dovevano nutrirsi dei cadaveri di piccoli dinosauri appena nati, di altre carcasse e probabilmente anche dei residui liquidi contenuti nell'uovo dopo la schiusa.

Sono stati trovati anche i resti di sauri predatori: i più comuni erano varanidi, i cui rappresentanti moderni sono stati osservati mentre scavavano e predavano uova di coccodrilli o di uccelli nidificanti sul terreno. Questi rettili erano probabilmente i maggiori predatori di uova di dinosauro non ancora schiuse. Un terzo predatore di cui vengono trovati sporadicamente i resti è *Troodon,* un piccolo dinosauro carnivoro che forse divorava i giovani ipsilofodonti.

Anche le uova di dinosauro disposte linearmente sono state trovate quasi soltanto sulla Egg Mountain; sebbene ne siano stati raccolti parecchi esemplari, non si è potuto stabilire quante covate rappresentassero a causa del modo inconsueto in cui erano distribuiti. Può darsi che lo schema lineare costituisse un carattere distintivo dei due ordini di dinosauri: gli ornitischi (prevalentemente erbivori) e i saurischi (prevalentemente carnivori). Gli ipsilofodonti, gli adrosauri, i ceratopsidi e altri dinosauri appartengono agli

Questa covata di 23 uova è stata asportata intatta da un nido di ipsilofodonte situato su una piccola altura denominata Egg Mountain, sempre nel sito della Willow Creek Anticline. Quelli visibili nella fotografia sono i fondi delle uova; il righello per il riferimento di scala è lungo cinque centimetri.

Questa rappresentazione della Egg Mountain mostra parecchi livelli contenenti nidi e resti scheletrici di dinosauri. Le zone in colore indicano l'estensione degli scavi, quelle in grigio aree o ancora coperte dai sedimenti o scomparse per erosione. I gruppi circolari di uova, situati nelle lievi depressioni, corrispondono a covate di ipsilofodonti. I piccoli ovali scuri sono uova di dinosauro di origine ignota. Gli scheletri con le lunghe zampe posteriori sono di ipsilofodonti; gli altri sono di lucertole, forse predatrici. La scala verticale è appositamente ingrandita.

In questa illustrazione la ricostruzione degli scheletri di un adrosauro e di un ipsilofodonte, ambedue adulti, è stata effettuata alla stessa scala. Mentre l'adrosauro raggiungeva probabilmente al più solo una lenta andatura dondolante l'ipsilofodonte doveva essere un corridore veloce.

In questa illustrazione, che è una litografia acquarellata di Henderson, sono rappresentati giovani adrosauri appena usciti dal nido, mentre pascolano accompagnati da un solo genitore, presumibilmente la madre. Probabilmente i piccoli avevano l'abitudine di cercare il cibo e di nutrirsi assieme ad altri componenti del branco. Dai resti rinvenuti nel sito della Willow Creek Anticline si può dedurre che gli adrosauri trascorressero la loro vita in gran parte o addirittura sempre riuniti in branchi oppure in altre forme di aggregazione sociale.

ornitischi, mentre i carnosauri e i brontosauri fanno parte dei saurischi.

Le uova attribuite al ceratopside *Protoceratops* (della Mongolia) si trovano in nidi circolari, al pari di quelle attribuite agli ipsilofodonti. Le prove materiali fornite dalla struttura del guscio e dalla forma del nido indicano che anche i nidi degli adrosauri erano circolari. Il brontosauro *Hypselosaurus,* un saurisco trovato in Francia, deponeva chiaramente le proprie uova in file. Non è escluso, pertanto, che gli ornitischi deponessero le uova in covate circolari e i saurischi in covate lineari. Forse le uova disposte linearmente trovate sulla Egg Mountain erano state deposte da *Troodon,* un piccolo saurisco predatore che avrebbe ricavato notevole vantaggio a riprodursi nella zona di nidificazione dei dinosauri non predatori.

Egg Island, come indica il nome, è completamente circondata da sedimenti lacustri. Anche se il lago non era probabilmente molto profondo e si prosciugò a intervalli nel periodo di tempo in cui l'isola venne utilizzata dai dinosauri per la nidificazione, esso fornì probabilmente un certo grado di protezione contro i predatori. Il sito è stato scoperto durante l'ultima settimana della stagione di ricerche del 1983 e quindi non si sono potuti effettuare scavi altrettanto estesi come sulla Egg Mountain. Ha comunque fornito già tre covate di uova di ipsilofodonti in due orizzonti separati. Una covata consiste in 19 uova non schiuse e ciascun uovo contiene i resti scheletrici di un embrione. Tutti gli scheletri sono delle stesse dimensioni e stanno a indicare che le uova sono state deposte nello stesso periodo e si sarebbero schiuse contemporaneamente.

Un altro sito della Willow Creek Anticline si sta dimostrando, sotto certi aspetti, ancora più ricco d'informazioni della Egg Mountain e della Egg Island. A circa un chilometro di distanza dalla Egg Mountain, in sedimenti alluvionali depositati anteriormente, sono stati trovati i resti di una zona di nidificazione di adrosauri. Per il momento sono stati scoperti otto nidi, i quali avrebbero ospitato uova pressappoco nello stesso periodo di tempo, dato che occupano un solo orizzonte sedimentario.

Gli adrosauri erano molto più grandi degli ipsilofodonti e gli adulti avevano mediamente una lunghezza di sette metri. Anche i loro nidi erano più ampi, con un diametro di circa due metri e una profondità di un metro. I nidi sarebbero stati sollevati sul terreno e probabilmente costruiti con fango; avevano una forma a catino e dovevano evidentemente contenere una covata circolare. Le dimensioni del nido e delle uova, tenuto conto anche del fatto che gli adrosauri avevano l'abitudine di disporre i loro nidi praticamente allo stesso modo degli ipsilofodonti, fanno ritenere che una covata comprendesse da 20 a 25 uova.

I nidi degli adrosauri, come quelli degli ipsilofodonti, erano fittamente raggruppati e separati l'uno dall'altro all'incirca della lunghezza di un adulto. Tuttavia essi presentano una differenza, in quanto non contengono fondi di guscio d'uovo intatti, ma abbondantissimi frammenti. Due nidi conservavano anche i resti di individui giovani. Uno di questi, danneggiato dalle intemperie, conteneva i corpi di sette piccoli, lunghi solamente 45 centimetri, mentre l'altro conteneva 15 corpi lunghi un metro circa. (Al momento della schiusa un adrosauro misurava da 30 a 35 centimetri di lunghezza.)

Gli adrosauri erano bipedi erbivori, privi degli abituali mezzi di difesa, salvo forse la capacità di scalciare e di calpestare con le zampe posteriori. Non avevano corna, placche, punte acuminate o epidermide corazzata. La struttura dello scheletro indica che la massima velocità di questi animali era quella che potevano raggiungere con una lenta andatura dondolante. Presumibilmente, i giovani dovevano essere in particolare molto vulnerabili nei confronti della predazione. Tuttavia, l'abbondanza di resti di adrosauro nei sedimenti della fine del Cretaceo sta a indicare che, fra tutti i gruppi di dinosauri, gli adrosauri sono quelli che hanno avuto il maggior successo. In che modo vi sono riusciti?

Le testimonianze fossili dimostrano che essi nidificavano in grosse, fitte colonie, sorvegliavano le proprie uova e, dopo la schiusa, curavano e nutrivano i piccoli. Come nel caso degli uccelli i cui piccoli nascono inetti, i giovani adrosauri rimanevano nel nido fino al momento in cui erano sufficientemente sviluppati per poter andare alla ricerca di cibo assieme agli adulti.

È stato anche accertato che gli adulti degli adrosauri erano in grado di lanciare segnali vocali di allarme per mettere in guardia altri membri di una colonia o di un branco: prove in questo senso sono contenute in lavori condotti indipendentemente da James A. Hopson dell'Università di Chicago, da David Weishampel della Florida International University e da me.

Queste supposizioni potrebbero non trovare favore tra tutti coloro che hanno studiato i dinosauri, ma pochi presumibilmente contesterebbero le ipotesi della nidificazione in colonie, della sorveglianza delle uova o delle manifestazioni vocali fra adulti, anche perché questi comportamenti si possono osservare anche nei coccodrilli attuali. Diversa è la questione della condivisione del cibo, un comportamento riscontrato prevalentemente negli uccelli e nei mammiferi. Molti pensano che ai dinosauri vengano attribuiti una fisiologia più complessa e un maggior numero di comportamenti di quanto sia possibile ammettere per un qualsiasi rettile estinto o vivente; essi probabilmente respingerebbero l'ipotesi della condivisione del cibo in quanto si tratta di un comportamento estremamente raro fra i rettili attuali.

La questione è di vedere se questi ultimi sono incapaci di sviluppare questo comportamento o altri o se, semplicemente, non ne hanno la necessità. Molti ritengono che i comportamenti fondamentali vengano trasmessi da una generazione all'altra attraverso i geni e che un animale, anche se non manifestasse un determinato comportamento fondamentale, sarebbe potenzialmente in grado di manifestarlo in caso di necessità. Per esempio, animali come i coccodrilli possono essere in grado di condividere il cibo, ma di fatto non vi sono obbligati. I coccodrilli sono predatori e aggressivi anche in giovane età. I dinosauri erbivori in giovanissima età, così come gli adrosauri, non solo erano chiaramente inoffensivi ma anche molto vulnerabili nei confronti di molti predatori, e pertanto i loro genitori erano costretti a procurare loro il cibo e ad assicurare loro un'efficace sorveglianza.

I sedimenti della Willow Creek Anticline forniscono anche le prove che gli adrosauri trascorrevano la maggior parte della loro esistenza, se non tutta, riuniti in branchi o in altre forme di aggregazione sociale; in sei distinte località, i resti indicano un'associazione fra adulti e giovani e il fatto che questi ultimi siano lunghi non più della metà dei primi fa pensare che individui subadulti (la cui lunghezza era compresa fra tre e sei metri) abbandonassero il gruppo per frequentare altri territori.

Sebbene queste interpretazioni siano basate su semplici congetture, esse sono corroborate dalle prove fornite dai siti di ritrovamento e dalla logica. Il fatto che nei nidi siano presenti solo i resti di giovani adrosauri delle stesse dimensioni sottintende una situazione ben diversa da qualunque cosa che potrebbe essere spiegata con il trasporto per mezzo di corsi d'acqua. Quegli individui sembrano morti sul posto e non esistono indizi di uccisioni da parte di predatori. Se i piccoli uccelli inetti vengono abbandonati dal genitore, o da ambedue i genitori, muoiono per denutrizione; d'altro canto sembra probabile che l'alimentazione si svolgesse all'interno del nido, dato che in caso contrario si sarebbe verificato un alto tasso di mortalità. Gli scheletri dei piccoli appena nati differiscono da quelli degli adulti solo per le dimensioni; si deve quindi ritenere che i giovani non potessero correre velocemente o eseguire altre manovre difensive con maggiore efficienza degli adulti. La sopravvivenza dei piccoli veniva probabilmente assicurata dal fatto che gli adulti li nutrivano dai bordi del nido; contribuiva inoltre alla loro sicurezza anche la fitta colonizzazione. La formazione dei branchi sarebbe stata una conseguenza del comportamento associativo degli adrosauri.

Per quanto riguarda gli ipsilofodonti della Egg Mountain, può darsi che, per loro, l'alimentazione all'interno del nido non fosse indispensabile. La morfologia dello scheletro indica che dovevano essere veloci corridori e la loro difesa primaria nei riguardi dei predatori consisteva probabilmente in una rapida fuga.

La coltivazione preistorica del riso nell'Asia sudorientale

Recenti ricerche archeologiche condotte nella regione dimostrano che la domesticazione di questo cereale condusse all'avvento di città stato e anche allo sviluppo di contatti commerciali con l'India e con la Cina

di C. F. W. Higham

Il riso è un alimento di cui si nutre più di metà del genere umano, eppure non sappiamo quasi nulla sulle fasi più antiche della sua domesticazione. Sugli inizi della coltivazione dell'orzo e del frumento nel Vicino Oriente, durante il Neolitico, e su quelli della coltivazione del granoturco nella Mesoamerica abbiamo una documentazione archeologica abbastanza buona, come pure l'abbiamo sul fatto che la domesticazione di queste piante permise, in entrambe le aree, una grande espansione demografica e l'avvento di stati urbani. Le informazioni analoghe sul riso sono state, invece, quanto meno scarse finché recenti ricerche nell'Asia sudorientale non cominciarono a rivelare almeno parte della storia.

In questo articolo mi propongo di riferire sullo stato presente delle conoscenze archeologiche in relazione alla coltivazione del riso nella regione. Una prima questione concerne il luogo, o i luoghi, in cui il riso divenne per la prima volta una componente fondamentale della dieta. Una seconda riguarda i processi che potrebbero aver contribuito a renderlo tale. Infine, se l'Asia sudorientale fu in effetti il sito iniziale dell'emergente coltivazione del riso, quali prove ci sono di una civiltà locale che sarebbe sorta di pari passo, come nel Vicino Oriente e nella Mesoamerica, prima che le grandi civiltà dell'India e della Cina segnassero la regione con la loro impronta verso la fine del I millennio a.C.?

È necessario innanzitutto fornire qualche indicazione sulla geografia e sulla botanica della regione e sul suo passato geologico recente. L'area è attraversata da tre fiumi principali: il Fiume Rosso, con delta nel Golfo del Tonchino, il Mekong, le cui molte foci dominano la Cambogia e il Vietnam meridionale, e il Chao Phraya o Menan, che trasforma Bangkok, la capitale della Thailandia, in una sorta di Venezia, prima di riversarsi nel Golfo del Siam. Riso allo stato selvatico cresce ancora oggi nelle tre valli di questi fiumi così come nelle valli dei loro molti affluenti. Nel vicino territorio della Cina la coltivazione del riso ebbe inizio, nella bassa valle dello Yangtze, attorno al 5000 a.C., cosicché è molto probabile che uno sviluppo simile abbia avuto luogo nelle valli fluviali dell'Asia sudorientale.

Importanti mutamenti nel livello del mare, associati alla fusione dei ghiacci del Pleistocene, provocarono l'inondazione di grandi aree della regione che erano in precedenza aride. Per esempio, circa 7000 anni fa il livello del mare era di vari metri superiore a quello attuale; il mare cominciò a ritirarsi al livello attuale solo fra 2000 e 3000 anni or sono. Così, i siti abitati dall'uomo che, in epoca glaciale, potevano essere localizzati sulla costa, si trovano oggi a una certa altezza al di sopra del mare, in terreni asciutti. Anche i siti archeologici localizzati sugli altipiani calcarei della regione sono rimasti indisturbati. Ciononostante le caverne nelle quali sono stati compiuti scavi, come la Grotta degli Spiriti nella Thailandia settentrionale (studiata da Chester Gorman dell'Università della Pennsylvania) e le caverne nella provincia nordvietnamita di Hoa Binh, offrono solo una visione parziale della vita preistorica della regione.

Le persone che cercarono rifugio nelle caverne sull'altopiano facevano parte per lo più di piccoli gruppi di cacciatori-raccoglitori. Il terreno è così propizio a un tale stile di vita che molte bande di cacciatori-raccoglitori vagarono nella regione sino a tempi storici avanzati, e alcune sono presenti in essa ancora oggi. Di fatto, però, la Grotta degli Spiriti è forse l'ultimo luogo nell'Asia sudorientale in cui si dovrebbero cercare tracce delle origini della coltivazione del riso. Essa si trova in un ambiente molto adatto per la caccia, ma è appollaiata su un pendio scosceso, molto lontana dall'acqua corrente. Benché il suolo della caverna sia stato setacciato accuratamente, Gorman non ha trovato in esso traccia di riso.

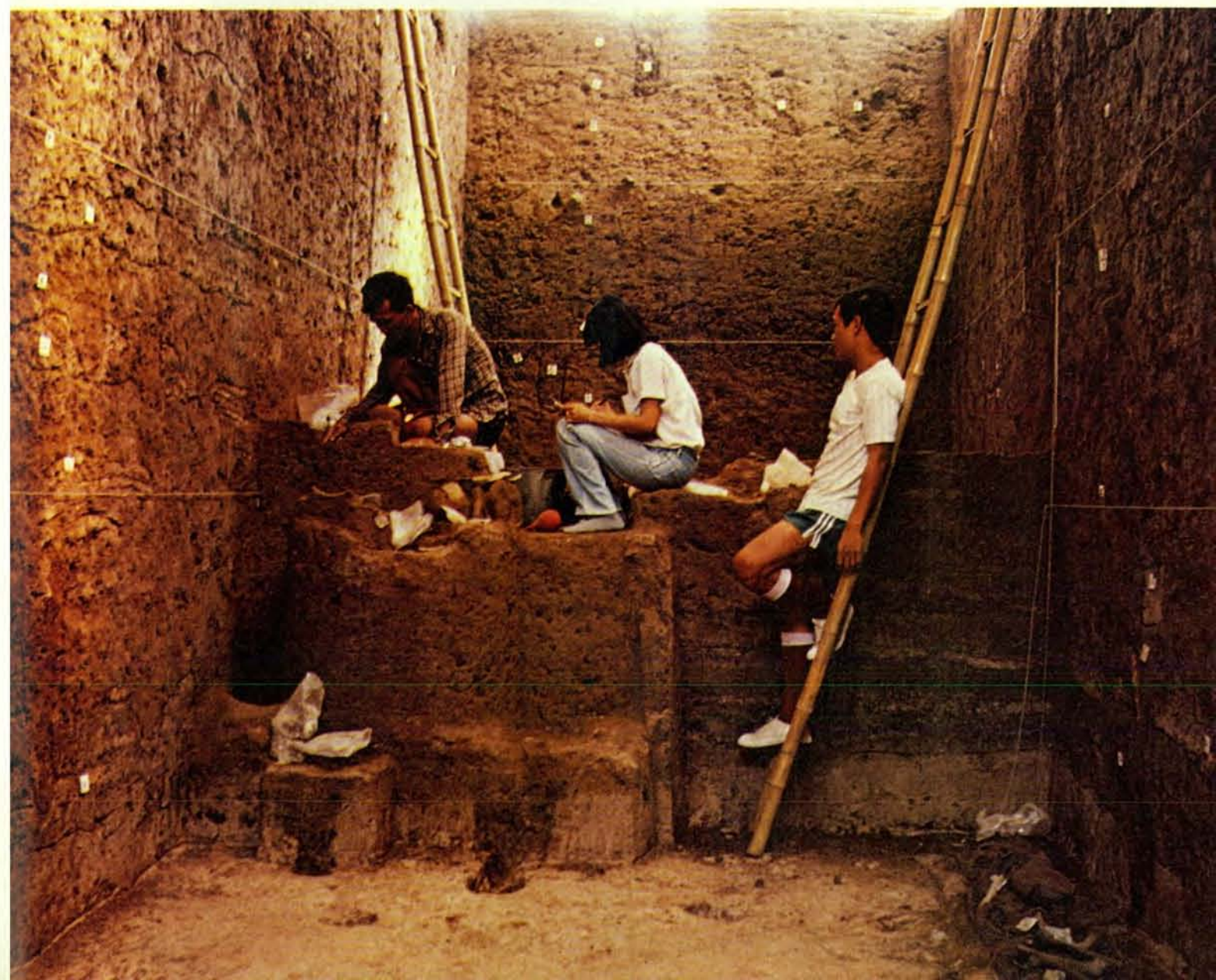
Diversa è la situazione nella striscia costiera ai piedi dell'altopiano di Hoa Binh. Scavi condotti in aree molto arretrate rispetto all'attuale linea di costa hanno riportato in luce numerosi ed estesi siti abitati dall'uomo sulla riva del mare, o in prossimità di essa, quando il livello del mare in epoca postglaciale era più alto. Ricognizioni hanno rivelato una situazione analoga ai margini del Golfo del Siam.

Il sito antico singolo più importante dell'area costiera thailandese, Khok Phanom Di, si trova oggi a più di 20 chilometri dalla costa. Nei secoli compresi tra il 5000 e il 2000 a.C., e forse per un periodo di tempo ancora più lungo, esso si trovava in una palude costiera di mangrovie o in prossimità di essa. Gli uomini che vivevano qui si cibavano di pesci, di molluschi e di crostacei adattati a vivere nelle acque salmastre fra le mangrovie. I resti dell'occupazione umana costituirono alla fine un cumulo di rifiuti alto più di 12 metri ed esteso per circa cinque ettari. All'interno di questo cumulo, in una sequenza stratigrafica chiaramente definita, si trovano tombe i cui occupanti furono cosparsi di ocra rossa prima della sepoltura; fra gli oggetti rinvenuti nelle tombe sono ornamenti personali, asce in pietra levigata e vasellame. Vasellame intero e cocci sono stati rinvenuti a tutti i livelli del sito e l'analisi dei cocci trovati a circa cinque metri al di sotto della superficie attuale dimostra che l'argilla era impastata con pula di riso.

Oggi è in corso di progettazione un'importante campagna di scavi a Khok Phanom Di, sotto il patrocinio del Dipartimento di Belle Arti del governo thailandese e dell'istituzione a cui io stesso appartengo, l'Università di Otago in Nuova Zelanda. Uno degli obiettivi di questa campagna è la verifica dell'ipotesi che gli abitanti di questo sito si siano dedicati a una coltivazione attiva del riso quando le acque del mare, ritirandosi, li lasciarono per così dire arenati, a una distanza piuttosto scomoda da quella fonte di cibo. L'uso della pula del riso mescolata ad argilla nella produzione di vasellame non è ovviamente una chiara prova della coltivazione del riso. La gente di Khok Phanom Di raccoglieva sicuramente riso selvatico, assieme ad altri cibi derivati da vegetali selvatici, per integrare la propria dieta.

Possiamo però essere certi di un fatto. Siano o no questi uomini diventati coltivatori di riso prima di abbandonare il sito attorno al 2000 a.C., a tale data o ancora prima il riso veniva coltivato altrove nell'Asia sudorientale, come dimostrano scavi recenti condotti sui due versanti delle montagne dell'Annam. Più di 50 piccoli insediamenti agricoli sono stati scoperti da archeologi vietnamiti sul versante orientale di quelle montagne, in quello che è noto come «paese di mezzo» a monte del delta del Fiume Rosso. Sul versante occidentale, nella media valle del Mekong, scavi archeologici condotti in tre siti thailandesi - Non Nok Tha, Ban Chiang e Ban Na Di - hanno rivelato che nell'area erano penetrati degli agricoltori che vi avevano stabilito i loro insediamenti.

In entrambe le aree citate i nuovi insediamenti si trovavano in luoghi che non presentano traccia di occupazione anteriore da parte di cacciatori-raccoglitori. I primi coltivatori scelsero per i loro villaggi un terreno leggermente elevato, località con facilità di accesso a terre basse, regolarmente inondate. Il terreno dal delta del Fiume Rosso verso l'interno era ideale per un simile insediamento. Numerose basse colline controllano un corso d'acqua che è un affluente della principale via d'acqua della regione. Gli scavi eseguiti in uno di questi siti del paese di mezzo, Phung Nguyen, dimostrano che i colonizzatori combinarono la coltivazione del riso con l'allevamento di animali e con la caccia. Essi erano anche eccellenti vasai e lavoravano con grande maestria la pietra. Il nome del sito è stato attribuito alla cultura di transizione del Neolitico, che si estese dal 2500 a.C. circa sino al 1800 a.C. sull'intero paese di mezzo. Solo gli strati più recenti di siti della cultura di Phung Nguyen contengono piccoli frammenti di metallo: si tratta di bronzo.

Oltre ai tre siti principali nella Thailandia nordorientale, ricognizioni recenti hanno localizzato numerosi siti di villaggi distribuiti secondo lo stesso modello di quelli della cultura di Phung Nguyen, ossia sul medio corso di piccoli affluenti del fiume principale. I più antichi fra questi insediamenti, fondati attorno al 2400 a.C., pare abbiano coperto solo un ettaro circa di terreno e difficilmente potrebbero avere avuto più di 50 o 100 occupanti. L'economia doveva essere impostata su basi molto simili a quella degli insediamenti dall'altra parte delle montagne. Il riso era coltivato in valli che venivano inondate con ritmo stagionale; si allevavano bovini, maiali e cani, e la dieta era integrata dai prodotti della caccia e della pesca. Poiché il clima della

Nella pagina a fronte si notano gli scavi di Ban Na Di, il sito di un villaggio nella Thailandia nordorientale, occupato per la prima volta attorno al 1400 a. C. I suoi abitanti coltivavano il riso e allevavano maiali e bovini. Erano anche abili vasai e lavoravano il bronzo; inoltre, seppellirono i morti nella stessa area sepolcrale per un periodo di un migliaio di anni prima di abbandonare il sito.

Questo bracciale di marmo, che fu rinvenuto nell'area sepolcrale di Ban Na Di, è uno dei vari ornamenti fatti con materiali importati da località lontane. Il suo valore viene dedotto dal fatto che, dopo essersi rotto, esso venne riparato con filo di bronzo, fatto passare attraverso coppie di fori.

Questa figurina di 15 centimetri, che rappresenta un toro, proviene da Ban Na Di ed è una delle molte statuine d'argilla, che rappresentano talvolta esseri umani, trovate nel terreno sepolcrale.

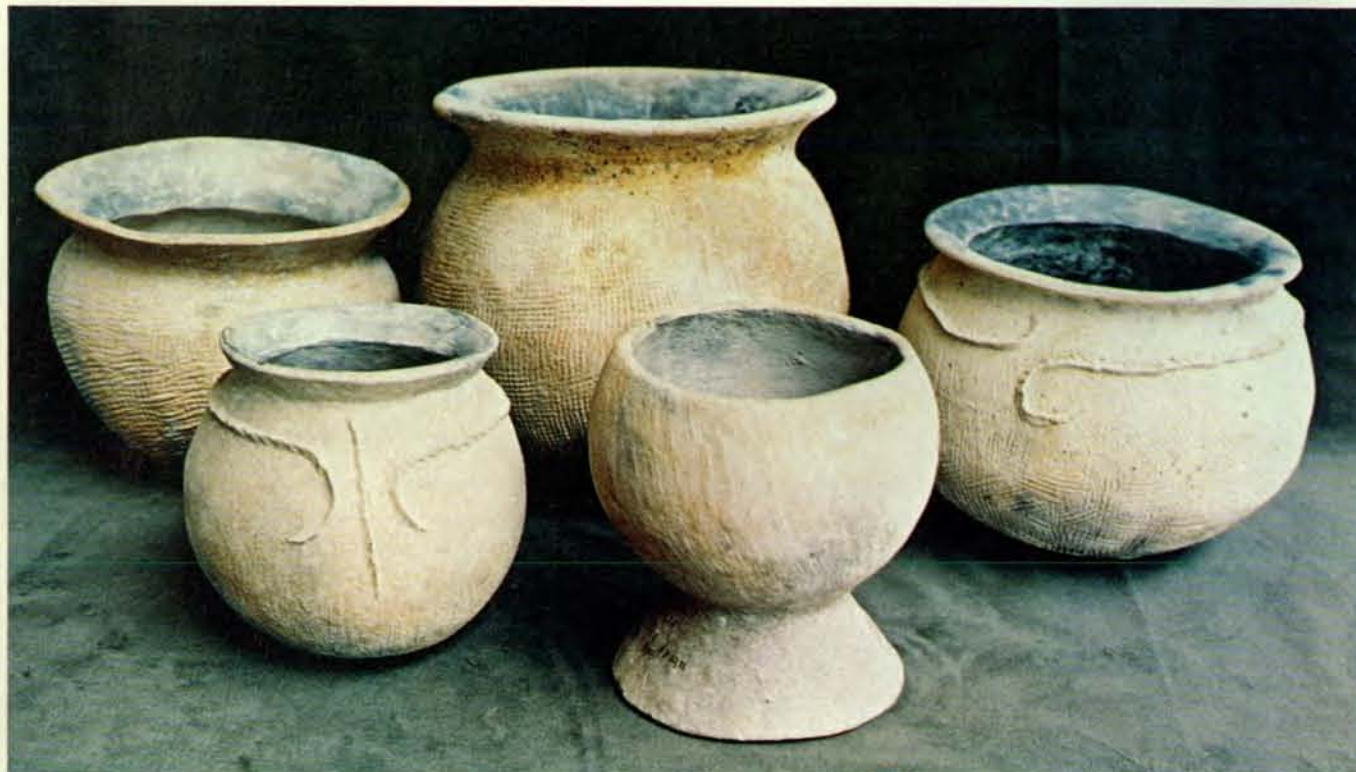
Thailandia nordorientale è caratterizzato da una siccità di sei mesi, interrotta dall'inizio dei monsoni, la sopravvivenza dei colonizzatori deve aver senz'altro richiesto l'immagazzinamento di gran parte del riso raccolto per consumarlo nella stagione secca. In ogni caso, come pare sia avvenuto in tutto il mondo, la sistemazione dei primi colonizzatori in insediamenti stabili dev'essere stata seguita da un incremento demografico, dalla crescita dei piccoli villaggi in insediamenti maggiori e dalla fondazione di altri insediamenti in una sorta di gemmazione continua.

Gli scavi eseguiti a Ban Na Di presentano una chiara visione del processo. Piccolo villaggio a una ventina di chilometri dal sito di un villaggio più grande, Ban Chiang, Ban Na Di fu fondato per la prima volta attorno al 1400 a.C., forse 1000 anni dopo il villaggio maggiore. La cultura materiale di Ban Na Di è così simile a quella di Ban Chiang, nonostante la data posteriore del sito più piccolo, da far supporre che il villaggio più recente sia stato fondato in conseguenza di un sovrappopolamento a Ban Chiang.

Ban Na Di non coprì mai più di tre ettari, e difficilmente la sua popolazione massima potrebbe aver superato le 150 persone. Oltre alla conoscenza della coltivazione del riso, i suoi fondatori portarono con sé la tecnologia della fusione del bronzo. Essi produssero principalmente asce e punte di freccia e sono state scoperte tracce di quasi tutte le fasi della lavorazione del metallo. La produzione di forme era naturalmente il primo passo necessario e quegli antichi lavoratori metallurgici le preparavano scolpendo una madreforma in due pezzi da un blocco di arenaria. Essi non colavano il bronzo fuso direttamente nella madreforma, ma riempivano quest'ultima con piombo, che ha un punto di fusione molto più basso di quello del bronzo, riducendo in tal modo il rischio che la madreforma si rompesse. I getti di piombo venivano poi rivestiti di argilla e, quando argilla e piombo venivano riscaldati assieme, l'argilla cuoceva e si induriva e il piombo fondeva e scorreva via. Le forme di argilla venivano poi riempite con una colata di bronzo, che l'artigiano preparava mettendo le proporzioni desiderate di rame e di stagno in un crogiolo di argilla, riscaldato in una fornace alimentata con carbone di legna e sottoposta a un tiraggio forzato. Nei livelli più antichi del sito di Ban Na Di troviamo le fornaci, i crogioli, frammenti di forme di arenaria e persino un frammento di uno dei getti di piombo.

Quasi tutti i manufatti in bronzo di Ban Na Di furono trovati però nella zona di sepoltura del villaggio. In effetti, sono proprio i resti umani e gli oggetti a essi associati a dirci di più sulla vita del villaggio. La stessa area fu usata come terreno sepolcrale quasi dalla fondazione di Ban Na Di e sino al suo abbandono attorno al 400 a.C. I morti venivano inumati, assieme ai loro ornamenti personali, in posizione supina con le gambe distese. Fra le suppellettili funerarie rinvenute sono vasi intatti, contenenti cibi come riso, pesce e,

Importanti siti dell'Asia sudorientale si distribuiscono per età dagli insediamenti collinari occupati da cacciatori-raccoglitori, come la Grotta degli Spiriti e l'antico cumulo di detriti di Khok Phanom Di, che risalgono a 7000 o più anni fa, a siti urbani del I millennio d.C., come Oc Eo e Angkor, che riflettono influenze indiane. L'area a ovest della moderna Hanoi è l'altopiano della provincia di Hoa Binh. I numerosi siti dei cacciatori-raccoglitori qui presenti hanno dato il nome di Hoabinhiano a questa fase della preistoria dell'Asia sudorientale, così come il sito dell'Età del bronzo di Dong S'on ha dato il nome ai molti siti dell'Età del bronzo scoperti recentemente nell'area del Fiume Rosso.

In questa tomba di Ban Na Di sono stati trovati vasi con tracce di cibi, di solito riso o pesce. In molti casi è stato trovato l'arto anteriore sinistro di un bue *(in primo piano a destra)*. L'importanza dell'individuo sepolto viene rilevata dal bracciale di bronzo all'avambraccio.

Cinque vasi d'argilla rinvenuti nel terreno sepolcrale di Ban Na Di, quattro dei quali con motivi ornamentali, contenevano un tempo cibi e bevande per il defunto. Il calice con piede è alto poco più di 20 centimetri. Le ultime sepolture nella necropoli risalgono al 400 a.C. circa.

in un caso, l'arto anteriore di un maialino. Alcuni recipienti che non contenevano cibi potrebbero aver contenuto acqua da bere o altri liquidi. È evidente che un'inumazione era l'occasione per il sacrificio di un animale; nella maggior parte dei casi accanto alla salma veniva deposto l'arto anteriore sinistro di un bue o di un maiale.

Molti fra gli oggetti decorativi - anelli per caviglie, bracciali e collane - erano fatti con materiali importati. Per esempio, un uomo fu sepolto con 14 bracciali di conchiglie. Le conchiglie appartengono al genere marino *Trochus*, un mollusco abitante in acque limpide che si trova oggi solo al largo della Thailandia, mille chilometri a sud di Ban Na Di. Altri braccialetti erano fatti di marmo, di travertino e di ardesia. Nessuno di questi materiali si trova nel luogo; la loro presenza, come quella delle conchiglie marine, indica l'esistenza di una consistente rete commerciale. I bracciali erano chiaramente molto pregiati. Per esempio, un bracciale di marmo che si era rotto era stato accuratamente riparato con filo di bronzo. Gli ornamenti che si deponevano accanto ai defunti non erano esclusivamente quelli che essi avevano portato da adulti: un altro bracciale di marmo era così piccolo che poteva essere stato infilato solo durante la fanciullezza.

Gli oggetti decorativi, comprese collane fatte con altre conchiglie marine, furono trovati tutti in sepolture concentrate in una parte del terreno sepolcrale di Ban Na Di. Molte fra le stesse tombe contenevano figurine di argilla raffiguranti bovini, cervi, elefanti e, in alcuni casi, esseri umani. Una tomba, in cui era stato sepolto un uomo, conteneva una «mandria» di sette figurine di bovini.

L'abbandono della zona di sepoltura di Ban Na Di, avvenuto circa mille anni dopo, non fu dovuto a un capriccio. Gli anni compresi fra il 500 a.C. e l'inizio dell'Era cristiana furono molto importanti nell'Asia sudorientale. L'analisi del vasellame rinvenuto nello strato di detriti accumulatisi al di sopra dell'area sepolcrale abbandonata ne indica l'appartenenza a una tradizione di ceramiche del tutto distinta rispetto al vasellame anteriore rinvenuto nello stesso sito. Un mutamento così brusco dopo dieci secoli di continuità documenta l'arrivo di una popolazione straniera.

I secoli finali del I millennio a.C. videro, almeno in certe aree dell'Asia sudorientale, una transizione dal modello anteriore di piccoli villaggi autonomi a un sistema sociale fondato sullo sviluppo di alcuni grandi centri e sull'avvento di capi. Questa transizione fu accompagnata (e senza dubbio anche accelerata) da crescenti contatti con le culture in espansione dell'India a ovest e della Cina a nord-est. I contatti ebbero luogo principalmente nelle regioni costiere; in effetti la più antica testimonianza oculare sulla società dell'Asia sudorientale è quella di un viaggiatore cinese, K'ang T'ai, che descrisse le popolazioni costiere della valle inferiore del Mekong attorno al 250 d.C. Egli riferì che vivevano in città cinte da mura contenenti palazzi oltre ad abitazioni più umili. Una parte dei servizi da tavola dei palazzi era d'argento e le autorità imponevano tasse sul commercio di argento, oro, perle e profumi.

Ricerche archeologiche condotte nella valle inferiore del Fiume Rosso a partire già dagli anni trenta hanno fornito informazioni su alcune di queste società più complesse dell'Asia sudorientale. A quell'epoca un ricercatore svedese, Olov R. F. Janse, scavò un certo numero di tombe in un ricco terreno sepolcrale sulla riva destra di un corso d'acqua vicino al delta del Fiume Rosso, in prossimità del piccolo villaggio di Dong S'on. La sua ricerca portò in luce, fra altri manufatti, pugnali, ornamenti per il corpo, tamburi e urne rituali, tutti in bronzo. Gli oggetti rinvenuti a Dong S'on erano così numerosi e ricordavano a tal punto le suppellettili nelle tombe dei capi, riportate in luce a quell'epoca nell'Europa orientale, che, per spiegare la presenza di questi manufatti di bronzo nel Vietnam, per qualche tempo si prestò seria considerazione a un'ipotetica migrazione «pontica» nell'Asia sudorientale.

Ricerche più recenti sulla cultura di Dong S'on ci narrano una storia del tutto diversa. Gli archeologi vietnamiti, che riconoscono nella popolazione di Dong S'on i loro diretti progenitori, ritengono che questo sito sia periferico rispetto al principale concentramento di insediamenti preistorici all'interno della regione del delta del Fiume Rosso. Inoltre essi sono ora in grado di ricondurre l'origine della cultura agli iniziali colonizzatori neolitici dei siti della cultura di Phung Nguyen, nel paese di mezzo, da me menzionati. Per esempio, alcuni fra i motivi presenti sui tamburi in bronzo trovati a Dong S'on sono incisi anche sul vasellame molto più antico della cultura di Phung Nguyen.

Le decorazioni dei tamburi indicano anche che quella di Dong S'on era una società guerriera ricca e aristocratica. Una scena comune raffigura uomini con acconciature complesse, armati di lance dalle aste infiocchettate. Un'altra scena, che corre attorno all'intero bordo dei tamburi, rappresenta flottiglie di canoe da guerra con guerrieri dalle acconciature ancora più complesse. Alcune canoe presentano in prossimità della poppa una piattaforma rialzata sulla quale sta in piedi un arciere. Al di sotto della piattaforma vi è quello che potrebbe essere un tamburo, usato presumibilmente per battere il tempo per i vogatori. Alte scene raffigurate sui tamburi presentano case su palafitte, suonatori di tamburi e altri musici.

Si è stimato che per la colata di un solo tamburo di Dong S'on fosse necessario fondere più di sette tonnellate di minerale di rame. Gli artigiani del bronzo producevano pugnali, spade e placche decorative per il corpo destinati agli aristocratici guerrieri. Fra gli oggetti da loro prodotti c'erano anche recipienti e mestoli in bronzo che probabilmente venivano usati nelle feste. Non tutta la produzione era riservata però alla guerra e al lusso; venne fabbricato anche un grande numero di pesanti vomeri in bronzo.

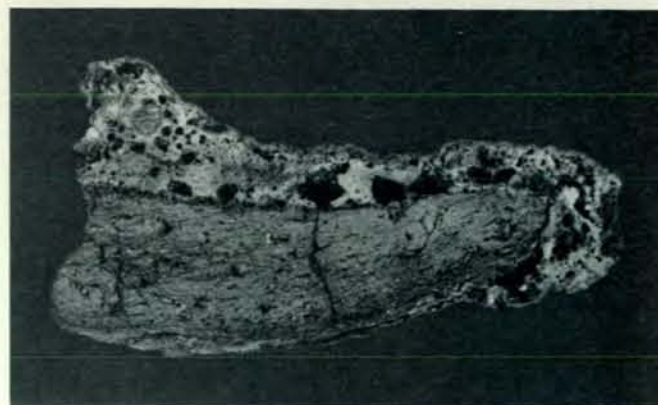
L'autorità degli aristocratici si riflette nelle loro suppellettili funerarie. Per esempio, nei due siti del Tonchino - Chau Can e Viet Khe - gli archeologi vietnamiti hanno scoperto sarcofagi in forma di barche. A Viet Khe, all'interno di un sarcofago lungo cinque metri vi erano più di cento manufatti in bronzo, comprendenti punte di lancia, asce, spade e tamburi.

Qual era la natura dell'economia di Dong S'on? Era sufficientemente produttiva da sostenere un urbanesimo maturo? Le scoperte archeologiche compiute nel Vietnam fanno pensare che si possa rispondere in modo affermativo alla seconda domanda. Il clima del delta del Fiume Rosso è unico nell'Asia sudorientale. Mentre la maggior parte di questa regione geografica ha una lunga stagione secca, i venti umidi della Cina meridionale portano nell'area del delta nubi basse e pioggerelle, rendendo possibili due raccolti di riso all'anno. Che i governanti di Dong S'on non abbiano trascurato questa possibilità è evidente sia dal gran numero di vomeri in bronzo trovati in vari siti di questa cultura sia dalla scoperta negli stessi siti di diverse varietà di riso, che indicano un doppio raccolto.

Qualche indizio di questo fatto si trova in resoconti cinesi della dinastia Han, quando i cinesi estesero la loro influenza in aree di

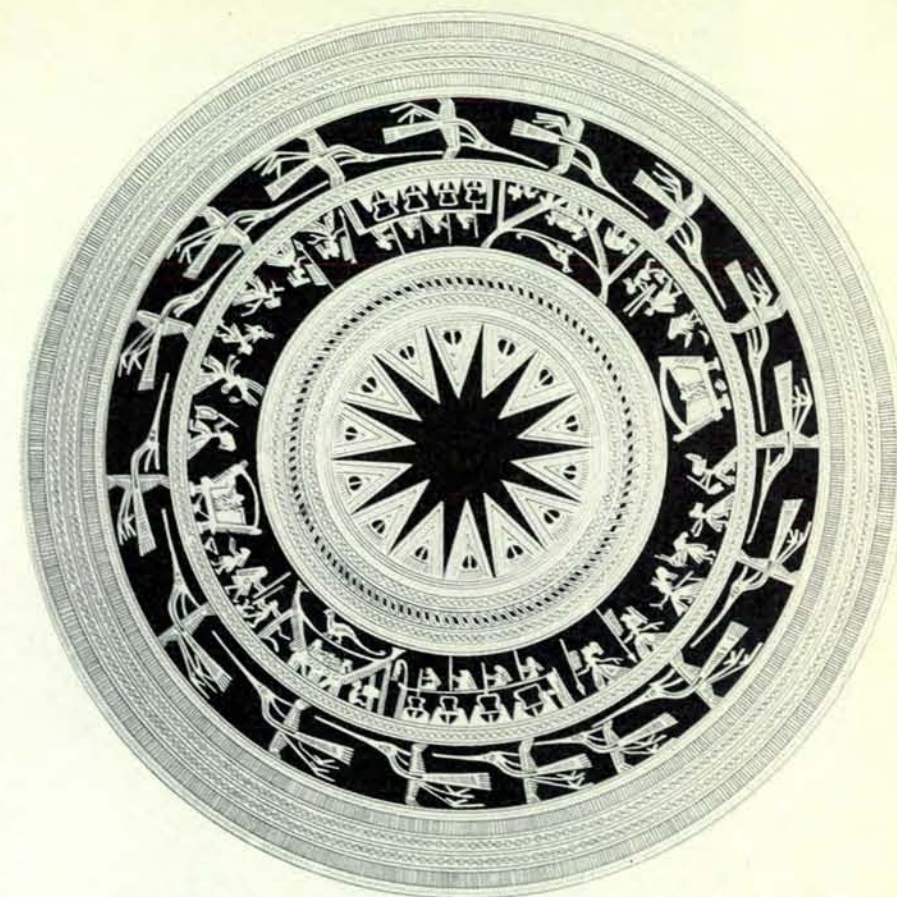
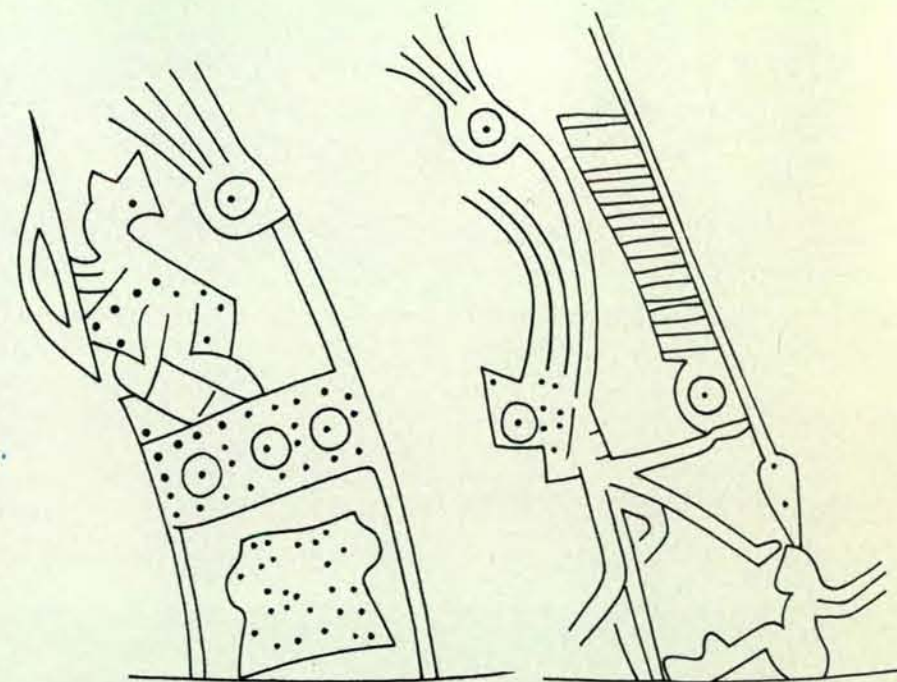
confine, occupando più o meno stabilmente territori controllati in precedenza da vietnamiti (Lac). In uno di tali resoconti si dice che i cinesi confermarono i diritti tradizionali dei locali capi Lac sulla terra del delta. Ricerche archeologiche recenti in un sito 15 chilometri a nord-ovest di Hanoi forniscono un indizio più tangibile: l'immenso sito di Co Loa, che fu per molto tempo oggetto di leggende.

Tutto ciò che rimane oggi di Co Loa sono tre cinte di bastioni in terra, alte circa sei metri e che si ergono imponenti al di sopra del paesaggio piatto del delta. I due bastioni esterni sono grosso modo ovali e quello interno rettangolare. Tutt'e tre sono circondati da fossati alimentati da un affluente del Fiume Rosso. L'area racchiusa dalla cinta più esterna ha un'estensione di almeno 500 ettari. Vasellame della cultura di Dong S'on è stato ritrovato sia al di sotto sia all'interno delle due cinte esterne mentre un tamburo in bronzo fu rinvenuto nell'area più interna. Esso conteneva più di cento vomeri. Dobbiamo apprendere molte altre cose sui modelli di insediamento e sull'organizzazione sociale della popolazione di Dong S'on in generale, ma è difficile non giungere alla conclusione che Co Loa fu una sede importante di autorità e forse addirittura il principale centro amministrativo della cultura di Dong S'on.

Il crogiolo di argilla a sinistra veniva riempito di rame e di stagno dagli artigiani di Ban Na Di e collocato in una buca rivestita di argilla e riempita di carbone di legna. Mediante un tiraggio d'aria forzato il carbone bruciava a una temperatura abbastanza alta per fondere i due metalli, ottenendo bronzo. I crogioli venivano spesso usati più di una volta, come indicano gli strati di residui su un frammento di uno di essi, a destra.

Questo frammento di arenaria fa parte di una forma usata a Ban Na Di per ottenere una replica in piombo di un oggetto di bronzo. Ricoperta la replica con argilla e fatta cuocere, il piombo fuso ne fuorusciva. La forma di argilla era così pronta per essere riempita con bronzo colato.

Qualunque sia stata la natura dell'organizzazione politica indigena nel delta del Fiume Rosso prima dell'Era cristiana e al suo inizio, essa era destinata ad avere breve vita. Nel 43 d.C. il generale della dinastia Han, Ma Yüan, annesse con la forza la regione e la trasformò in una prefettura di confine della dinastia regnante. All'esterno dell'area governata direttamente dai cinesi vi sono però prove archeologiche crescenti dello sviluppo nell'Asia sudorientale prima di società gerarchizzate e poi di stati.

Alcuni esempi del delta del Mekong sono sufficienti a suggerire questa tendenza. Innanzitutto, fotografie aeree recenti hanno rivelato l'esistenza nel delta di una rete di antichi canali. Ricognizioni al suolo suggeriscono che essi risalgono al III secolo d.C., l'epoca in cui K'ang T'ai visitò la regione. In secondo luogo, da mezzo secolo o più sono note (e saccheggiate alla ricerca di oro) le rovine di un porto nel delta, Oc Eo, mentre negli anni quaranta sull'area di 450 ettari l'archeologo francese Louis Malleret condusse una ricerca sistematica. Egli trovò che il sito era stato un centro di produzione, fra l'altro, di perline di vetro e di amuleti di stagno fuso. Fra gli oggetti riportati in luce da Malleret, che attestano contatti con i vicini stati indiani, sono sigilli recanti incise iscrizioni in alfabeto *brāhmī*. Che il sito fosse inserito in una vasta rete commerciale è confermato ulteriormente dalla scoperta a Oc Eo di monete romane del II secolo d.C.

Le rovine della civiltà khmer in Cambogia, una civiltà influenzata dal buddhismo, come quella di Angkor, sono troppo note perché sia il caso di soffermarvisi qui. Molto meno noti sono i numerosi siti di città circondate da fossati nel delta del Chao Phraya, associate alla prima importante organizzazione politica della Thailandia, il regno di Dvaravati, su cui solo oggi si comincia ad acquisire qualche informazione in più. Un sito del regno di Dvaravati si trova a soli pochi chilometri dall'antico cumulo di detriti costiero di Khok Phanom Di, dove il nostro progetto di ricerca potrebbe rivelare se i cacciatori-raccoglitori 5000 anni fa coltivavano riso o si limitavano a raccogliere il cereale allo stato selvatico. I siti archeologici della regione ci ricordano che lo sviluppo preistorico della coltivazione del riso fu alla base degli stati sia antichi sia moderni dell'Asia sudorientale.

Questa parte di tamburo in bronzo, scoperta meno di un anno fa a Co Loa, un sito urbano su un affluente del Fiume Rosso, comprende nella sua complessa decorazione due bande di figure. La banda esterna raffigura uccelli volanti identificati come gru o aironi. La banda interna presenta una scena ripetuta due volte. Nella parte bassa dell'immagine si nota una serie di quattro tamburi di dimensione crescente e su di essi sono seduti quattro suonatori di tamburo. Alla loro destra vi sono due uomini dall'acconciatura elaborata, armati di lance dalle aste infiocchettate, i quali guidano una processione di altre quattro persone, forse musici. Più a destra, al di là di quella che potrebbe essere un'abitazione, due individui stanno pestando in un mortaio, forse mondano del riso. Segue poi una struttura che sembra un rifugio per animali, dopo di che si ripete la stessa scena.

La decorazione di tamburo, che rappresenta una canoa da guerra della cultura di Dong S'on, comprende tra l'altro le figure riportate nel disegno. Si notano un arciere in piedi su una piattaforma sopraelevata *(a sinistra)* e un uomo armato di lancia, con acconciatura elaborata, che sta minacciando un prigioniero seduto *(a destra)*. Sotto la piattaforma si osserva un tamburo.

(RI)CREAZIONI AL CALCOLATORE

di Fred Gruenberger

Come e perché trattare numeri con migliaia di cifre

[Nota redazionale: L'autore delle «(Ri)Creazioni al calcolatore» di questo mese, Fred Gruenberger, insegna scienze del calcolatore alla California State University di Northridge. Ha cominciato a occuparsi di macchine da calcolo più di 40 anni fa; da allora ha pubblicato 28 libri e tra il 1977 e il 1981 è stato direttore della rivista «Popular Computing».

A partire dal mese prossimo, la rubrica sarà affidata ad A. K. Dewdney, professore associato di scienze del calcolatore alla University of Western Ontario. Gli interessi professionali di Dewdney vanno soprattutto alla matematica del discreto e alla teoria della computazione, ma è noto a molti lettori per le sue ricerche scientifiche e tecniche sul «planiverso» bidimensionale, descritto a suo tempo da Martin Gardner nella rubrica «Giochi matematici» («Le Scienze», n. 147, novembre 1980). Dewdney stesso ha scritto un resoconto dei propri lavori (*The Planiverse: Computer Contact with a Two-dimensional World*) che la Poseidon Press ha pubblicato di recente; una raccolta di saggi su argomenti vari di scienza del calcolatore, *The Turing Omnibus*, sarà pubblicata l'anno prossimo dalla Computer Science Press.]

Se avete una calcolatrice con un tasto per elevare al quadrato un numero, provate a fare questa operazione: battete il numero 1,0000001 e premete 27 volte il tasto del quadrato. L'operazione equivale a elevare il numero iniziale alla 134217728-esima potenza. Il risultato corretto, con una precisione di dieci cifre significative, è 674 530,4707, ma la calcolatrice darà quasi sicuramente un risultato diverso. Il problema ha lo scopo di evidenziare il grado di precisione della macchina. La tabella di pagina 122 dà i risultati ottenuti con varie calcolatrici e con versioni dei linguaggi di programmazione BASIC e Fortran che girano su alcuni calcolatori. Nessuna di queste macchine raggiunge le sette cifre esatte.

Nella maggior parte delle calcolatrici elettroniche, l'operazione di elevazione al quadrato non equivale all'immissione di un numero e alla sua moltiplicazione per se stesso. In quest'operazione, i fattori sono limitati per dimensione o per precisione dal numero delle cifre che la macchina può visualizzare. L'elevazione al quadrato, invece, agisce sulla rappresentazione di un numero immagazzinato nella macchina, che di solito include alcune «cifre di guardia», cioè cifre extra che entrano in ogni calcolo ma sono nascoste all'operatore. Così, se calcolate la radice quadrata di 2 su una macchina con un visualizzatore di otto cifre, il risultato apparirà come 1,4142136, ma sarà immagazzinato all'interno nella forma 1,414213562. Premendo il tasto del quadrato si dovrebbe tornare al valore 2,0000000, mentre la moltiplicazione di 1,4142136 per se stesso dà 2,000001 06.

Nella maggior parte dei calcoli, un errore nella settima posizione decimale ha scarse conseguenze. Supponiamo, però, che il calcolo sia inserito in un programma di calcolatore che eseguirà una certa sequenza di istruzioni se il valore è esattamente 2, mentre ne eseguirà un'altra in caso contrario; l'effetto di una imprecisione potrebbe essere catastrofico. Il modo più sicuro per evitare questo pericolo sta nell'arrotondare il valore calcolato a un livello di precisione noto prima di controllare se è uguale a 2. In altre circostanze lo stratagemma dell'arrotondamento non è di alcun aiuto. Nel problema consistente nell'elevare ripetutamente al quadrato una frazione decimale, l'unico modo per migliorare la qualità del risultato è mantenere una maggior precisione lungo tutti gli stadi del calcolo.

Per la maggior parte dei lavori numerici, un livello di precisione di otto o nove cifre è ampio. Dovrebbe essere sufficiente per tenere i conti di un libretto di assegni, se le cifre in gioco non sono superiori a un miliardo di lire circa. Non si conosce alcuna costante in natura con una precisione maggiore di 12 cifre significative. Per raggiungere una precisione nel risultato anche solo modesta, però, può essere necessaria una precisione molto maggiore nel corso del calcolo. Quando si eleva 27 volte al quadrato 1,0000001, per avere 10 cifre corrette tutti i calcoli vanno effettuati con una precisione di 15 cifre.

ELEVAZIONI AL QUADRATO	POTENZA	OTTO CIFRE DI PRECISIONE	15 CIFRE DI PRECISIONE
0	1	1,0000001	1,00000010000000
1	2	1,0000002	1,00000020000001
2	4	1,0000004	1,00000040000006
3	8	1,0000008	1,00000080000028
4	16	1,0000016	1,00000160000120
5	32	1,0000032	1,00000320000496
6	64	1,0000064	1,00000640002016
7	128	1,0000128	1,00001280008128
8	256	1,0000256	1,00002560032640
9	512	1,0000512	1,00005120130818
10	1024	1,0001024	1,00010240523794
11	2048	1,0002048	1,00020482096271
12	4096	1,0004096	1,00040968387705
13	8192	1,0008192	1,00081953559497
14	16384	1,0016391	1,00163974282853
15	32768	1,0032809	1,00328217441361
16	65536	1,0065726	1,00657512149610
17	131072	1,0131884	1,01319347521490
18	262144	1,0265507	1,02656101821804
19	524288	1,0538063	1,05382752412486
20	1048576	1,1105077	1,11055245060312
21	2097152	1,2332274	1,23332674554061
22	4194304	1,5208498	1,52109486126578
23	8388608	2,3129841	2,31372957696917
24	16777216	5,3498954	5,35334455534193
25	33554432	28,621381	28,6582979282091
26	67108864	819,18345	821,298040141993
27	134217728	671061,52	674530,470741078

Risultato della ripetuta elevazione al quadrato di 1,0000001 con precisione limitata

Un esempio di calcolo in cui è necessaria un'assoluta precisione è la ricerca continua di numeri primi sempre più grandi. Per molti anni, prima dello sviluppo del calcolatore elettronico digitale, il maggiore numero primo conosciuto era $2^{127} - 1$, che ha 39 cifre nella notazione decimale. Con l'aiuto del calcolatore, a partire dal 1952, il primato è stato infranto 16 volte; attualmente spetta a un numero ($2^{132\,049} - 1$) di 39 751 cifre. Nella verifica della «primalità» (cioè nella procedura per stabilire se un numero x può essere diviso per qualche numero diverso da 1 e da x stesso) tutte le operazioni aritmetiche devono essere esatte.

Calcolare il valore di π è un altro annoso problema che richiede altissima precisione. Il primato è stato detenuto a lungo dal matematico inglese William Shanks che, lavorando con carta e matita, ha calcolato 528 cifre corrette (e altre 179 sbagliate). Oggi π è noto fino a 8 388 608 cifre.

Dato che la meccanizzazione dell'aritmetica ad alta precisione è iniziata più di 30 anni fa e che da allora la potenza di calcolo è divenuta sempre meno costosa e più ampiamente disponibile, ci si potrebbe aspettare che l'elenco dei risultati noti sia piuttosto esteso. In realtà, l'elenco dei numeri noti con elevata precisione è decisamente corto.

Conosciamo la radice quadrata di 2 fino a un milione di posti decimali e fino a 1000 posti la radice cubica di 16.

La radice reale dell'equazione di Wallis, $X^3 - 2X - 5 = 0$, è conosciuta fino a 4000 cifre. (L'equazione è quella scelta dallo scienziato inglese del XVII secolo John Wallis per un'illustrazione del metodo di Newton per la soluzione numerica di equazioni; è servita poi come verifica di altri metodi di approssimazione.)

Harry L. Nelson, del Lawrence Livermore National Laboratory (che un tempo detenne insieme al collega David Slowinski il primato per il più grande numero primo conosciuto) ha calcolato il fattoriale di un milione, vale a dire il prodotto $1\,000\,000 \times 999\,999 \times \ldots \times 2 \times 1$. Il risultato ha 5 565 709 cifre e riempie un pacco di carta comune da stampante alto 13 centimetri.

Un problema noto come il problema del 196 è stato sviluppato per 50 000 stadi del calcolo, e a quel punto i numeri in gioco sono lunghi 21 000 cifre. Un altro problema tormentoso, cioè il problema del $3N + 1$, è stato studiato per valori isolati con 1000 cifre. Più avanti, esaminerò entrambi i problemi più in dettaglio.

La successione di Fibonacci (1, 1, 2, 3, 5, 8, 13, 21 ..., in cui ogni termine è uguale alla somma dei due termini precedenti) è stata esplicitamente calcolata per i primi 10 000 termini. Al 10 000-esimo termine i numeri hanno più di 2000 cifre.

Il numero di Eulero, e, la base dei logaritmi naturali, è stato calcolato per più di 125 000 posti decimali.

Si potrebbe citare qualche altro risultato isolato. Per esempio, R. William Gosper della Symbolics, Inc., ha calcolato, con un metodo basato sulla manipolazione di frazioni continue, 2800 cifre della radice settima di 306. (La radice non ha significato particolare; il compito era stato scelto a caso per verificare il metodo.)

1,0000001

1,00000020000001

1,0000004000000600000040000001

1,000000800000280000056000007000000560000028000000080000001

1,000001600001200000560000182000043680008008001144000128700011440000800800
04368000182000005600000120000001600000001

1,000003200004960004960003596002013760906192336585705183028048806451225290
245025792874737364714356565722780108044657227671435634737362257928529024486
4512242804880105183003365856090619202013760035960000496000004960000032000
0001

1,000006400020160041664063537676245194974430121663461681220599659328917561
1024560167423597795591072087154275016633545248630060887595709345161126819 9
15383185691423735290107259571090772323598587085564911661593733567974201161
43434357277218808573355568943751110599510580696377806180968267449354037803
99083454638968579161289551024136215303148046482901272501914064547774155796
97132175700584954616543012161994974368762451206353760041664000201600000640
000001

1,000012800081280341377066802645669423620652593817045586724716639752509356
37066049668878686323398085866658787740848661692888859950737447654753345168
00296374007374699571055600724167009930092944010785440807255491125231778992
84998080130939659275749411332443815476335984490898028690821180944687760178
26651186922820900475103999094028004541163740697539733888428367675593254423
04791447891813575210242928814336165345917104081421677640174634634397085972
83900954070660781524156253596638345132414037094689364780030653669917579320
39263495546002578420990239188092834549068962510444861858929119375002542434
57232088976910936243623532598515862472240337396250214803336301073293071392
40881177571510121069678262892429678985166947033041357598928186137247128344
63748230478358549260032564540002953869332914454490582230113562493732890458
57150156087070217497026618525795742361122645664010668000034137600081280000
1280000001

Risultati esatti per alcune elevazioni al quadrato di 1,0000001

La maggior parte dei problemi elencati sono di un tipo che richiede intrinsecamente un'alta precisione; sono ardui e stimolanti proprio perché bisogna tener conto di un numero di cifre superiore a quanto la maggior parte della gente possa mai immaginare. Un problema d'alta precisione più pratico e comune, e meno artificiale, è quello che sorge nella soluzione delle radici di un'equazione quadratica (un'equazione della forma $AX^2 + BX + C = 0$). La quantità $B^2 - 4AC$, il cosiddetto «discriminante» dell'equazione, stabilisce se le radici sono reali o immaginarie. Se il discriminante è positivo, le radici sono reali; se è negativo, le radici sono immaginarie; se $B^2 - 4AC$ è esattamente uguale a zero, l'equazione ha due radici uguali. Anche il più piccolo errore nella valutazione del discriminante, quindi, può portare a una differenza qualitativa nella soluzione dell'equazione.

Una analoga sensibilità a piccoli errori numerici può aversi nella soluzione di un sistema di equazioni lineari simultanee. Si consideri il sistema

$$53{,}17X - 18{,}91Y - 5{,}67Z = -174{,}65$$
$$-12{,}65X + 36{,}16Y - 47{,}08Z = 298{,}59$$
$$303{,}80X - 203{,}03Y + 112{,}89Z = -1769{,}02$$

Il sistema è stato costruito in modo che le tre equazioni siano soddisfatte dai valori $X = -3$, $Y = 2$ e $Z = -4$, ma questi valori non rappresentano l'unica soluzione. In effetti, anche se la cosa non risulta evidente a prima vista, non c'è una soluzione unica perché due delle equazioni descrivono piani paralleli l'uno all'altro. (La terza equazione è uguale a cinque volte la prima meno tre volte la seconda e non contribuisce quindi con alcuna informazione indipendente.)

L'impossibilità di risolvere le equazioni può essere scoperta calcolando la grandezza detta determinante della matrice di coefficienti. Il determinante di una matrice si valuta formando tutte le possibili combinazioni di elementi che non sono né nella stessa colonna né nella stessa riga; gli elementi di ogni combinazione vengono moltiplicati, poi vengono sommati i prodotti. Se il determinante risulta pari a zero, il sistema di equazioni non ha alcuna soluzione e si sa di non dover procedere oltre (o, più importante, si può predisporre che un programma per calcolatore si arresti a questo punto). In questo caso, però, un calcolo del determinante secondo un metodo standard che utilizzi un'aritmetica con una precisione di nove cifre dà una risposta pari non a zero, bensì a $-0{,}000202179$. Altri metodi danno forse un risultato corretto in questo esempio ma non in altri; il fatto è che con una precisione di nove cifre il risultato non è affidabile. Per di più, il sistema di equazioni qui presentato è piccolo, con coefficienti di non più di cinque cifre significative; quando il sistema è più grande, il problema diventa grave.

Molti problemi di teoria dei numeri, e di altri campi della matematica, richiedono una precisione estrema per la rappresentazione di interi molto grandi. Ne è un esempio il problema del 196. Per affrontare il problema, iniziate con un qualsiasi intero positivo di due o più cifre. Invertite l'ordine delle cifre e sommate il numero invertito a quello originale; ora invertite la somma e sommate di nuovo, poi continuate il processo finché non ottenete un numero palindromo: un numero uguale sia che lo si legga da sinistra sia che lo si legga da destra *(si veda l'illustra-*

CALCOLATRICE TASCABILE	RISULTATO	ERRORE PER PERCENTUALE
Texas Instruments SR-52	674520.6053	0.00146
Hewlett-Packard 33, 67, 41C	674494.0561	0.00540
Sharp Electronics EL 506	674492.75	0.00559
Monroe Calculator 1930	674383.1672	0.02183
Texas Instruments 30	674363.69	0.02473
CALCOLATORE E LINGUAGGIO		
Fortran con doppia precisione (CDC Cyber)	674530.5363	0.00000973
Fortran a otto cifre (CDC Cyber)	674530.5765	0.00001568
Apple II BASIC	22723.9709	96.63114
IBM Personal Computer BASIC	8850273	1212.06423
Ontel BASIC	8886690	1217.46401

Precisione di alcune macchine e linguaggi di programmazione nell'elevazione al quadrato ripetuta

zione in basso in questa pagina). Per la maggior parte dei numeri di partenza, si raggiunge rapidamente un palindromo; la serie che inizia con 195, per esempio, termina dopo appena quattro passi. Il più piccolo numero che sembra non diventare mai palindromo con questa procedura è 196; come detto in precedenza, si è fatta una verifica lunga 50 000 passi. Tra i primi 10 000 interi, ce ne sono 5996 che apparentemente non generano un palindromo per quanto a lungo si applichi la procedura (anche se questa congettura non è stata confermata).

Il problema del $3N + 1$ è stato presentato in questa rubrica nel fascicolo dello scorso marzo. Iniziate con un qualsiasi intero positivo N; se N è dispari, sostituitelo con $3N + 1$; se N è pari, sostituitelo con $N/2$. Continuate finché N è uguale a 1. Per esempio, quando il valore iniziale di N è 9, il processo comporta 20 termini: 9, 28, 14, 7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1. Questa semplice procedura conduce a numerosi misteri. La sequenza termina sempre con 1, indipendentemente dal valore iniziale di N scelto? Esiste una regolarità nel numero di passi richiesti? Esiste, cioè, una formula basata sul valore di N che permetta di prevedere il numero di passi necessari? Per qualsiasi numero di passi scelto, c'è sempre un valore dispari di N che generi una sequenza di quella lunghezza? (È facile trovare un valore pari di N con qualsiasi lunghezza di percorso data: corrisponde a 2 elevato a una potenza pari alla lunghezza del processo meno 1.)

Nel 1980 ipotizzai che il numero medio di termini per la convergenza nel problema del $3N + 1$ fosse approssimativamente $24{,}64D - 101$, dove D è il numero di cifre del valore iniziale di N. La stima era basata su calcoli fatti con numeri che arrivavano a circa 200 cifre, che allora ritenevo abbastanza grandi. Con un programma per aritmetica ad alta precisione sono riuscito a verificare la congettura per alcuni valori più grossi. Per un numero di 1000 cifre, la formula prevede la convergenza in 25 539 passi. Scoprii che quando N è uguale a 1 ... (998 zeri) ... 1, ossia $10^{1000} + 1$, la serie scende a 1 in 23 069 passi. Il numero 55 ... (997 zeri) ... 1 porta a una serie con 24 413 termini. Sulla base di queste prove preliminari, quindi, sembra che la velocità media di convergenza sia stabile e prevedibile.

Un altro fenomeno che si osserva all'inizio con piccoli valori di N, e che sembra perdurare con valori più grandi, è una certa tendenza a generare, per molti valori consecutivi di N, serie della stessa lunghezza. Successioni di questo genere, anzi, paiono aumentare col crescere di N. Per esempio, i 230 interi consecutivi a partire da

912345678912345678900-
000000000000000000001

generano tutti una serie con 997 termini. (Si noti che la formula empirica prevede per questo numero 959 termini.)

193	194	195	196	197	198	199	200
391	491	591	691	791	891	991	002
1124	685	786	887	988	1089	1190	202
4211	586	687	788	889	9801	0911	
5335	1271	1473	1675	1877	10890	2101	
	1721	3741	5761	7781	09801	1012	
	2992	5214	7436	9658	20691	3113	
		4125	6347	8569	19602		
		9339	13783	18227	40293		
			38731	72281	39204		
			52514	90508	79497		
			41525	80509			
			94309	171017			
			90349	710171			
			187088	881188			
			880781				
			1067869				
			9687601				
			10755470				
			07455701				
			18211171				
			17111281				
			35322452				
			25422353				
			60744805				
			50844706				
			111589511				
			⋮				

Valutazione del problema del 196 per alcuni valori iniziali

Le operazioni aritmetiche che possono essere effettuate direttamente dall'unità centrale di elaborazione di un normale microcalcolatore offrono una precisione molto limitata. In molte macchine, le uniche operazioni fornite esplicitamente sono la somma e la sottrazione di interi di lunghezza non superiore a 16 bit, ossia cifre binarie; nella notazione decimale, il massimo numero rappresentabile è 65 536, quindi il livello di precisione è inferiore a cinque cifre decimali. Anche con i calcolatori più potenti, si può ottenere una precisione superiore a qualche decina di cifre solo con un programma che combini molte operazioni elementari per suddividere un numero grande in «pezzi» più piccoli.

I diagrammi di flusso di pagina 124 mostrano un algoritmo per un particolare calcolo ad alta precisione: la valutazione del massimo numero primo conosciuto, $2^{132049} - 1$. La procedura è stata creata avendo in mente un particolare microelaboratore, vale a dire il 6502, prodotto dalla MOS Technology, che si trova in calcolatori della Apple Computer Inc., della Commodore Business Machines

Inc. e di altre aziende. È facile adattare l'algoritmo ad altri microelaboratori.

Forse la decisione fondamentale nella progettazione di un programma del genere è la scelta del modo in cui un numero debba essere rappresentato nella memoria del calcolatore. L'elaboratore può operare solo su valori binari, ma ovviamente si preferirebbe che il risultato del calcolo sia mostrato in forma decimale. Un utile compromesso è lo schema detto decimale codificato in binario, in cui ogni cifra decimale è rappresentata separatamente dal suo valore binario equivalente. Il 6502 organizza la memoria di un calcolatore in «byte», unità di otto bit, e pertanto è comodo immagazzinare un numero con una cifra decimale per byte. (Non è il metodo più efficace, ma è quello più semplice.)

Il primo passo dell'algoritmo consiste nel pulire un'area della memoria di 39 760 byte, ponendo ogni byte uguale a zero. Si pone poi un valore iniziale pari a 1 nell'area di lavoro pulita, in modo che l'area contenga 39 759 zeri seguiti da un solo 1. Si dà un valore iniziale pari a zero a un contatore, necessario per seguire il procedere del calcolo.

Diagrammi di flusso per il calcolo del numero primo di 39 751 cifre $2^{132\,049} - 1$.

La sezione principale del programma è un'iterazione che richiama ripetutamente un sottoprogramma avente la funzione di raddoppiare il numero immagazzinato nello spazio di lavoro. Dopo ogni raddoppio il contatore è aumentato di 1 e il suo valore è confrontato con 132 049. Se non si è raggiunto quel limite, viene richiamato ancora il sottoprogramma di raddoppio; quando si raggiunge il limite, il programma esce dall'iterazione. Infine, si sottrae 1 al valore nell'area di lavoro e si visualizza il risultato.

Nel diagramma di flusso a sinistra, è segnato a colori il sottoprogramma di raddoppio; nel diagramma di flusso a destra appaiono le istruzioni che compongono il sottoprogramma. Ogni volta che si richiama il programma, si pone un indice *(X)* nell'indirizzo più basso dello spazio di lavoro, dove è immagazzinata la cifra più a destra, quella meno significativa, del numero. Il valore immagazzinato a quell'indirizzo viene poi raddoppiato sommandolo a se stesso. Come quando si calcolano somme a mano, la complicazione sorge soprattutto quando si deve fare un «riporto» da una posizione all'altra col procedere della somma. Se una somma precedente ha generato una cifra di riporto, questa deve essere aggiunta al nuovo risultato. Questo risultato, a sua volta, deve essere controllato per una eventuale cifra di riporto: se la somma è maggiore di 9, deve essere aggiustata sottraendo 10 e la cifra di riporto va posta uguale a 1. Il procedimento viene ripetuto per tutti i 39 760 byte dell'area di lavoro.

Il microelaboratore 6502 dell'Apple II opera a una velocità superiore alle 250 000 operazioni al secondo. Ciononostante, la procedura mostrata nello schema a blocchi richiederebbe 120 ore per calcolare la 132 049-esima potenza di 2. Piccole scorciatoie molto semplici potrebbero ridurre di molto il tempo di elaborazione. Per esempio, non è necessario raddoppiare tutte le cifre dell'area di lavoro nei primi stadi del calcolo, quando sono quasi tutte uguali a zero. Lo spazio di lavoro potrebbe iniziare con una lunghezza di, diciamo, 150 byte e crescere di tre byte ogni 10 potenze. Questa strategia richiederebbe di per sé l'esecuzione di alcune istruzioni aggiuntive, ma l'effetto globale sarebbe una maggiore velocità.

Tutti i numeri che intervengono nel calcolo di $2^{132\,049} - 1$ sono interi, ma molti problemi si affrontano meglio esprimendo le quantità in notazione scientifica, in cui un numero è formato da una frazione decimale detta mantissa e da un esponente che dà la grandezza, ossia la potenza di 10. Per esempio, nel numero che rappresenta l'anno in corso la mantissa è 1,984, l'esponente è + 3 e il numero completo è scritto $1{,}984 \times 10^3$. Un programma ad alta precisione per trattare numeri di questo genere è nccessariamente più complicato di uno che abbia a che fare con interi perché ogni numero ha diverse parti (non solo la mantissa e l'esponente, ma anche i loro segni).

La maggior parte dei linguaggi di programmazione ad alto livello danno la possibilità di calcoli in notazione scientifica: si parla spesso in questo caso di «aritmetica in virgola mobile». I numeri inferiori a una certa dimensione appaiono come normali frazioni decimali, ma i valori sono immagazzinati all'interno come una mantissa e un esponente. Lo spazio assegnato ai vari elementi del numero determina la precisione e il dominio di valori che si possono rappresentare. Dando più spazio alla mantissa si aumenta la precisione; un esponente più elevato offre un dominio maggiore. La versione del BASIC incorporata nell'Apple II ha un livello di precisione di circa nove cifre.

Per avere una precisione che sia sensibilmente superiore è nuovamente necessario ricorrere a una soluzione software. Herman P. Robinson, un tempo al Lawrence Livermore Laboratory, ha scritto un pacchetto di programmi ad alta precisione in notazione scientifica nel linguaggio macchina del microelaboratore 6502. I programmi possono operare a qualsiasi livello di precisione fino a 600 cifre decimali e consentono di avere esponenti fino a 9999. Questi limiti sono stati scelti perché corrispondono a certe caratteristiche dell'elaboratore. Una mantissa di 600 cifre, un esponente di quattro cifre e i loro segni possono stare tutti in 256 byte; nel 6502, un blocco di 256 byte è una «pagina» di memoria.

Nei programmi di Robinson, le operazioni che si possono effettuare sui numeri comprendono le operazioni aritmetiche elementari, il logaritmo, l'esponenziale, la radice quadrata e varie funzioni trigonometriche. Vi sono anche funzioni meno comuni, come le trasformazioni di Eulero e di van Wijngaarden per sommare serie a lenta convergenza. Nel pacchetto sono registrati i valori di qualcosa come 26 costanti e 8000 numeri primi. Può essere usato come una calcolatrice da tavolo ad alta precisione, oppure le funzioni possono essere richiamate entro un programma. Le versioni preliminari indicano che le operazioni aritmetiche e le funzioni sono precise fino al limite di 600 cifre.

Per coloro che si interessano di calcoli numerici, una prima conoscenza col calcolatore digitale è a volte scoraggiante: dato che i calcoli elementari della macchina sono veloci e fondamentalmente impeccabili, l'aspettaiva ingenua è che si possano fare con grande facilità elaborate analisi numeriche. Si rimane quindi delusi nell'apprendere che la radice quinta di 100 (la quantità che gli astronomi indicano come un ordine di grandezza) non può essere determinata con una precisione molto maggiore di quella che fornisce una calcolatrice tascabile. Un pacchetto di programmi come quello di Robinson salva alcune delle promesse del calcolatore: la radice quinta di 100 può essere calcolata fino a 100 posti decimali nel giro di pochi minuti.